Supplementary Information

Site-specific Covalent Capture of Human O^6-Alkylguanine-DNA-alkyltransferase Using Single-stranded Intrastrand Cross-linked DNA

Derek K. O'Flaherty and Christopher J. Wilds[§]*

Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke St. West, Montréal, Québec (Canada) H4B 1R6

Contents **Page**

Supplementary Discussion

Synthesis of nucleoside dimers required for IaCL assembly **S-4**

Supplementary Methods

Synthesis and characterization of nucleosides and oligonucleotides **S-5**
UV thermal denaturation **S-20**
Circular dichroism (CD) spectroscopy **S-20**
Molecular modelling **S-20**
Protein expression and purification **S-21**
AGT repair assay of IaCL DNA duplexes **S-21**
Identification of IaCL Repair Product by SDS-PAGE **S-21**
Identification of repair product (DPC) from hAGT-mediated repair of IaCL DNA **S-22**

Supplementary Schemes

Supplementary Scheme 1 - Synthesis of O^4-dT-alkylene-O^4-dT dimer phosphoramidites **7a** and **7b**. **S-23**
Supplementary Scheme 2 - Synthesis of O^6-dG-alkylene-O^4-dT dimer phosphoramidites **11a** and **11b**. **S-24**
Supplementary Scheme 3 - Synthesis of O^4-dG-alkylene-O^4-dT dimer phosphoramidites **16a** and **16b**. **S-25**

Supplementary Figures

Supplementary Figure 1 - 500 MHz ^{1}H NMR spectrum of compound (**2**) (in $CDCl_3$) **S-26**
Supplementary Figure 2 - 125.7 MHz ^{13}C NMR spectrum of compound (**2**) (in $CDCl_3$) **S-27**
Supplementary Figure 3 - 500 MHz ^{1}H NMR spectrum of compound (**5a**) (in $CDCl_3$) **S-28**
Supplementary Figure 4 - 125.7 MHz ^{13}C NMR spectrum of compound (**5a**) (in $CDCl_3$) **S-29**

Supplementary Figure 5 - 500 MHz ^{1}H NMR spectrum of compound (**5b**) (in $CDCl_3$) **S-30**
Supplementary Figure 6 - 125.7 MHz ^{13}C NMR spectrum of compound (**5b**) (in $CDCl_3$) **S-31**
Supplementary Figure 7 - 500 MHz ^{1}H NMR spectrum of compound (**6a**) (in $CDCl_3$) **S-32**
Supplementary Figure 8 - 125.7 MHz ^{13}C NMR spectrum of compound (**6a**) (in $CDCl_3$) **S-33**
Supplementary Figure 9 - 500 MHz ^{1}H NMR spectrum of compound (**6b**) (in $CDCl_3$) **S-34**
Supplementary Figure 10 - 125.7 MHz ^{13}C NMR spectrum of compound (**6b**) (in $CDCl_3$) **S-35**
Supplementary Figure 11 - 500 MHz ^{1}H NMR spectrum of compound (**7a**) (in d_6-acetone) **S-36**
Supplementary Figure 12 - 125.7 MHz ^{13}C NMR spectrum of compound (**7a**) (in d_6-acetone) **S-37**
Supplementary Figure 13 - 202.3 MHz ^{31}P NMR spectrum of compound (**7a**) (in d_6-acetone) **S-38**
Supplementary Figure 14 - 500 MHz ^{1}H NMR spectrum of compound (**7b**) (in d_6-acetone) **S-39**
Supplementary Figure 15 - 125.7 MHz ^{13}C NMR spectrum of compound (**7b**) (in d_6-acetone) **S-40**
Supplementary Figure 16 - 202.3 MHz ^{31}P NMR spectrum of compound (**7b**) (in d_6-acetone) **S-41**
Supplementary Figure 17 - 500 MHz ^{1}H NMR spectrum of compound (**8a**) (in $CDCl_3$) **S-42**
Supplementary Figure 18 - 125.7 MHz ^{13}C NMR spectrum of compound (**8a**) (in $CDCl_3$) **S-43**
Supplementary Figure 19 - 500 MHz ^{1}H NMR spectrum of compound (**8b**) (in $CDCl_3$) **S-44**
Supplementary Figure 20 - 125.7 MHz ^{13}C NMR spectrum of compound (**8b**) (in $CDCl_3$) **S-45**
Supplementary Figure 21 - 500 MHz ^{1}H NMR spectrum of compound (**9a**) (in $CDCl_3$) **S-46**
Supplementary Figure 22 - 125.7 MHz ^{13}C NMR spectrum of compound (**9a**) (in $CDCl_3$) **S-47**
Supplementary Figure 23 - 500 MHz ^{1}H NMR spectrum of compound (**9b**) (in $CDCl_3$) **S-48**
Supplementary Figure 24 - 125.7 MHz ^{13}C NMR spectrum of compound (**9b**) (in $CDCl_3$) **S-49**
Supplementary Figure 25 - 500 MHz ^{1}H NMR spectrum of compound (**10a**) (in $CDCl_3$) **S-50**
Supplementary Figure 26 - 125.7 MHz ^{13}C NMR spectrum of compound (**10a**) (in $CDCl_3$) **S-51**
Supplementary Figure 27 - 500 MHz ^{1}H NMR spectrum of compound (**10b**) (in $CDCl_3$) **S-52**
Supplementary Figure 28 - 125.7 MHz ^{13}C NMR spectrum of compound (**10b**) (in $CDCl_3$) **S-53**
Supplementary Figure 29 - 500 MHz ^{1}H NMR spectrum of compound (**11a**) (in d_6-acetone) **S-54**
Supplementary Figure 30 - 125.7 MHz ^{13}C NMR spectrum of compound (**11a**) (in d_6-acetone) **S-55**
Supplementary Figure 31 - 202.3 MHz ^{31}P NMR spectrum of compound (**11a**) (in d_6-acetone) **S-56**
Supplementary Figure 32 - 500 MHz ^{1}H NMR spectrum of compound (**11b**) (in d_6-acetone) **S-57**
Supplementary Figure 33 - 125.7 MHz ^{13}C NMR spectrum of compound (**11b**) (in d_6-acetone) **S-58**
Supplementary Figure 34 - 202.3 MHz ^{31}P NMR spectrum of compound (**11b**) (in d_6-acetone) **S-59**
Supplementary Figure 35 - 500 MHz ^{1}H NMR spectrum of compound (**13**) (in $CDCl_3$) **S-60**
Supplementary Figure 36 - 125.7 MHz ^{13}C NMR spectrum of compound (**13**) (in $CDCl_3$) **S-61**
Supplementary Figure 37 - 500 MHz ^{1}H NMR spectrum of compound (**14a**) (in $CDCl_3$) **S-62**
Supplementary Figure 38 - 125.7 MHz ^{13}C NMR spectrum of compound (**14a**) (in $CDCl_3$) **S-63**
Supplementary Figure 39 - 500 MHz ^{1}H NMR spectrum of compound (**14b**) (in $CDCl_3$) **S-64**
Supplementary Figure 40 - 125.7 MHz ^{13}C NMR spectrum of compound (**14b**) (in $CDCl_3$) **S-65**
Supplementary Figure 41 - 500 MHz ^{1}H NMR spectrum of compound (**15a**) (in $CDCl_3$) **S-66**
Supplementary Figure 42 - 125.7 MHz ^{13}C NMR spectrum of compound (**15a**) (in $CDCl_3$) **S-67**
Supplementary Figure 43 - 500 MHz ^{1}H NMR spectrum of compound (**15b**) (in $CDCl_3$) **S-68**
Supplementary Figure 44 - 125.7 MHz ^{13}C NMR spectrum of compound (**15b**) (in $CDCl_3$) **S-69**
Supplementary Figure 45 - 500 MHz ^{1}H NMR spectrum of compound (**16a**) (in d_6-acetone) **S-70**
Supplementary Figure 46 - 125.7 MHz ^{13}C NMR spectrum of compound (**16a**) (in d_6-acetone) **S-71**
Supplementary Figure 47 - 202.3 MHz ^{31}P NMR spectrum of compound (**16a**) (in d_6-acetone) **S-72**
Supplementary Figure 48 - 500 MHz ^{1}H NMR spectrum of compound (**16b**) (in d_6-acetone) **S-73**
Supplementary Figure 49 - 125.7 MHz ^{13}C NMR spectrum of compound (**16b**) (in d_6-acetone) **S-74**
Supplementary Figure 50 - 202.3 MHz ^{31}P NMR spectrum of compound (**16b**) (in d_6-acetone) **S-75**
Supplementary Figure 51 - ESI MS spectrum of oligonucleotide **TT4** **S-76**
Supplementary Figure 52 - ESI MS spectrum of oligonucleotide **TT7** **S-77**

Supplementary Figure 53 -	ESI MS spectrum of oligonucleotide **GT4**	**S-78**
Supplementary Figure 54 -	ESI MS spectrum of oligonucleotide **GT7**	**S-79**
Supplementary Figure 55 -	ESI MS spectrum of oligonucleotide **TG4**	**S-80**
Supplementary Figure 56 -	ESI MS spectrum of oligonucleotide **TG7**	**S-81**
Supplementary Figure 57 -	UV thermal denaturation experiments for unmodified DNA duplex (**TT**), **TT4**, and **TT7**	**S-82**
Supplementary Figure 58 -	UV thermal denaturation experiments for unmodified DNA duplex (**GT**), **GT4**, and **GT7**	**S-82**
Supplementary Figure 59 -	UV thermal denaturation experiments for unmodified DNA duplex (**TG**), **TG4**, and **TG7**	**S-83**
Supplementary Figure 60 -	Circular dichroism spectra for unmodified DNA duplex (**TT**), **TT4**, **TT7**	**S-83**
Supplementary Figure 61 -	Circular dichroism spectra for unmodified DNA duplex (**GT**), **GT4**, **GT7**	**S-84**
Supplementary Figure 62 -	Molecular models of unmodified control duplex (**TT**) and duplexes containing **TT4** and **TT7** that were geometry optimized using the AMBER forcefield	**S-85**
Supplementary Figure 63 -	Molecular models of unmodified control duplex (**GT**) and duplexes containing **GT4** and **GT7** that were geometry optimized using the AMBER forcefield	**S-86**
Supplementary Figure 64 -	Molecular models of unmodified control duplex (**TG**) and duplexes containing **TG4** and **TG7** that were geometry optimized using the AMBER forcefield	**S-87**
Supplementary Figure 65 -	Total repair assay of (**A**) **TT4** and (**B**) **TT7** by hAGT, OGT, Ada-C, and OGT S134P	**S-88**
Supplementary Figure 66 -	Total repair assay of (**A**) **GT4** and (**B**) **GT7** by hAGT, OGT, Ada-C, and OGT S134P	**S-89**
Supplementary Figure 67 -	Time course repair gel of duplexes containing (**A**) **TG4** or (**B**) **TG7** by hAGT	**S-90**
Supplementary Figure 68 -	Time course repair gel of duplexes containing **TG7** by 30-fold protein of (**A**) Ada-C or (**B**) hAGT.	**S-91**
Supplementary Figure 69 -	15% SDS-PAGE of repair reaction between hAGT (300 pmol) and IaCL DNA (300 pmol) for either 1h or 3.5 h	**S-92**
Supplementary Figure 70 -	15% SDS-PAGE of repair reaction between hAGT (300 pmol) and IaCL DNA (300 pmol) for 1h at 37^{o}C. Panel (**A**) is Coomassie Blue stained. Panel (**B**) gel visua Fluorescence monitoring of the fluorescein tag.	**S-92**
Supplementary References		**S-93**

Supplementary Discussion

Synthesis of nucleoside dimers required for IaCL assembly

The preparation of the modified DNA in scales and purity amenable for biochemical studies was accomplished by a combination of small molecule and automated solid-phase synthesis. The preparation of cross-linked phosphoramidites **7a** and **7b**, required for the preparation of **TT4** and **TT7** respectively, is shown in **Supplementary Scheme 1**. The synthesis was adapted from procedures developed by Swann.[1,2] The introduction of an additional protecting group at the 3'-hydroxyl position was necessary to prepare **TT4** and **TT7**. The levulinyl protecting group was chosen based on its compatibility with various other nucleoside constructs.[3] Protected monomer **2** was prepared by selectively introducing the TBS group at the 5'-hydroxyl of commercially available thymidine (**1**), followed by esterifcation of the 3'-hydroxyl with levulinic acid, in good yield over two steps. The preparation of the convertible nucleoside required optimization since the previously utilized procedure (for 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidine) led to formation of side-product, presumably occurring at the carbonyl groups found in the levulinyl moiety. The production of the desired convertible nucleoside was accomplished by the addition of $POCl_3$ in two portions (1.5 equivalences followed by another 1.1 after 5 min of stirring on ice). The reaction was quenched after 20-30 min to minimize side-product formation and, once worked-up, the intermediate was used as is without further purification for the subsequent step. Dimerization was achieved in the presence of DBU and a slight excess of the corresponding alcohol **4a** and **4b**, which have been synthesized previously by our group.[2] Removal of the levulinyl group was achieved using a standard protocol,[3] with the slight modifications, to furnish intermediates **6a** and **6b**.

The synthesis of cross-linked phosphoramidites **11a** and **11b**, required for the synthesis of **GT4** and **GT7**, respectively, is shown in **Supplementary Scheme 2** and began with the preparation of hydroxybutyl and hydroxyheptyl mono-adducts **8a** and **8b**, respectively, from intermediate protected monomer **2**. Only a slight excess of the corresponding diol was utilized to minimize side-product formation. Compound **8a** or **8b** was coupled with a protected dG analogue[4] *via* the Mitsunobu reaction to produce dimers **9a** or **9b**. The Mitsunobu reaction was selected on the basis that it can selectively introduce alkyl groups at the O^6-atom of dG.[4–8] Removal of the levulinyl groups was achieved as described above to produce dimers **10a** and **10b**.

Phosphoramidites **16a** and **16b**, required for the preparation of **TG4** and **TG7**, were prepared in a similar fashion as described for **11a** and **11b** and shown in **Supplementary Scheme 3**. Commercially available *N*2-phenoxyacetyl-2'-deoxyguanosine was converted to the fully protected monomer **13** using an analogous procedure to the thymidinyl intermediate **2**. Dimers **14a** or **14b** were prepared using the Mitsunobu coupling of **4a** or **4b** with monomer **13**. Finally, levulinyl deprotection of **14a** and **14b** furnished the corresponding alcohols **15a** and **15b**, respectively.

Supplementary methods

Synthesis and characterization of nucleosides and oligonucleotides

General

*N*2-phenoxyacetyl-2'-deoxyguanosine , 5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine, 5'-*O*-(4,4'-dimethoxytrityl)-thymidine, thymidine, and N,N-diisopropylaminocyanoethylphosphonamidic chloride were purchased from ChemGenes, Inc. (Wilmington, MA). 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine and was prepared according to previously reported procedures.[4] Mono-adduct **4a** and **4b** were prepared accordingly to previous published procedures.[2] 5'-*O*-Dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-N,N-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside–CPG supports were purchased from Glen Research (Sterling, Virginia). 6-FAM phosphoramidites and 6-FAM lcaa CPG (DMT) 500Å were purchased from ChemGenes (Wilmington, MA) and used according to the manufacturer's protocols. All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) purchased from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using an 7T-LTQ FT ICR instrument (Thermo Scientific), at the Concordia University Centre for Structural and Functional Genomics. The mass spectrometer was operated in full scan, positive ion detection mode. ESI mass spectra for oligonucleotides were obtained at the CBAMS using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. Ampicillin, isopropyl β-D-thiogalactopyranoside (IPTG), and most other biochemical reagents as well as polyacrylamide gel materials were purchased from Bioshop Canada Inc (Burlington, ON). Ni-NTA Superflow Resin was purchased from Qiagen (Mississauga, ON). Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC) Nitro- cellulose filters (0.20 µm) were obtained from Millipore. XL-10 Gold and Bl21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). T4 polynucleotide kinase (PNK) was obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). Phusion Polymerase was obtained from New England Biolabs (Ipswich, MA). DNA primers for site directed mutagenesis and cloning were purchased from Biocorp (Montreal, QC).

Chemical synthesis of nucleosides

The synthesis of phosphoramidites **7a** and **7b** are shown in **Supplementary Scheme 1**

5'-O-tert-butyldimethylsilyl-3'-O-levulinyl- thymidine (**2**)

Thymidine (2.0 g, 8.3 mmol) was co-evaporated in vacuo with anhydrous pyridine (2 x 5 mL). To a solution of dried thymidine in pyridine (50 mL) at 0°C was added TBS-Cl (1.3 g, 8.6 mmol) and allowed to stir. After 3h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL) forming a white precipitate in suspension. The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL) and once with brine (50 mL), which removed the precipitate, followed by drying over anhydrous Na_2SO_4 (~ 4 g). The solvent was removed in vacuo and analysis by TLC revealed that the crude product was of sufficient purity to proceed to the subsequent step. To a solution of the 5'-*O* mono-silylated thymidine adduct (~2.9 g, ~8.1 mmol), EDC (3.1 g, 16 mmol) and DMAP (cat) in dioxane (64 mL) was added levulinic acid (1.9 g, 16 mmol) and allowed to stir. After 4h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL), washed with aqueous $NaHCO_3$ (3% w/v, 2 x 50 mL) and once with brine (50 mL), and dried over anhydrous Na_2SO_4 (~ 4 g). The solvent was removed in vacuo and the product was purified *via* flash column chromatography using MeOH : DCM (1% to 2.5%) as eluent to afford 3.2 g (85% over two steps) of a slightly beige powder. R_f (SiO_2 TLC): 0.53 MeOH : DCM (5%). $\lambda_{max(MeCN)}$ = 264 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 9.38 (s, 1H, NH), 7.54 (d, 1H, H6, *J* = 1 Hz), 6.33 (dd, 1H, H1', *J* = 9, 6 Hz), 5.24 (m, 1H, H3'), 4.08 (m, 1H, H4'), 2.75 (m, 2H, CH_2CO), 2.58 (m, 2H, $OOCCH_2$), 2.39 (m, 1H, H2'), 2.18 (s, 3H, $COCH_3$), 2.08 (m, 1H, H2''), 1.90 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.91 (s, 9H, $SiC(CH_3)_3$), 0.109 (s, 3H, $SiCH_3$), 0.105 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 205.7, 176.8, 172.4, 168.4, 164.0, 150.5, 135.2, 111.1, 85.4, 84.7, 75.6, 63.6, 38.0, 37.7, 29.7, 27.9, 25.9, 18.3, 12.4, , -5.42, -5.52. IR (thin film); ν_{max} (cm^{-1}) = 3066, 2955, 2929, 2856, 1715, 1471, 1276, 1257, 1073, 1005, 837. HRMS (ESI-MS) *m/z* calculated for $C_{21}H_{35}N_2O_7Si^+$ 455.2208: found 455.2211 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O- (levulinyl)-thymidinyl]}-butane (**5a**)

To a solution of compound **2** (1.2 g, 2.6 mmol) , triazole (4.0 g, 58 mmol), NEt_3 (8.3 mL, 59 mmol) in MeCN : DCM (32mL, 1/1 v/v) stirring at 0°C for 10 min was added $POCl_3$ dropwise (0.36 mL, 3.9 mmol). After 5 min, another portion of $POCl_3$ (0.27 mL, 2.9 mmol) was added which resulted in the formation of a precipitate. After 20 min, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL). The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL) and once with brine (50 mL), followed by drying over anhydrous Na_2SO_4 (~ 4 g). The solvent was removed in vacuo and crude intermediate **3** was used directly for the subsequent dimerization reaction. To a solution of **3** (~ 1.0 g, ~2.1 mmol) and **4a** (1.7 g, 2.3 mmol) in MeCN (15 mL) was added DBU (0.92 mL, 6.2 mmol) and allowed to stir at 21 °C. After 20h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL). The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated in vacuo. The crude product was purified *via* flash column chromatography using EtOAc : hexanes (7 :3 to 100:0) as eluent to afford 1.2 g of **5a** (52 %) as a colorless foam. R_f (SiO_2 TLC): 0.35 EtOAc . $\lambda_{max(MeCN)}$ =283 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.94 (s, 1H, H6), 7.77

(s, 1H, H6), 7.42-7.40 (m, 2H Ar), 7.31-7.23 (m, 7H, Ar), 6.84-6.81 (m, 4H, Ar), 6.41 (dd, 1H, H1'a, *J* = 9, 5.5 Hz), 6.32 (dd, 1H, H1'b, *J* = 6 Hz), 5.24 (m, 1H, H3'a), 4.48-4.43 (m, 5H, H3'b & 2 x $ArOCH_2$), 4.13 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.91-3.90 (m, 2H, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 2.75 (m, 2H, CH_2CO), 2.64-2.47 (m, 4H, H2'a, H2'b & $OOCCH_2$), 2.23-2.18 (m, 4H, H2''b & $COCH_3$), 2.04-1.92 (m, 4H, H2''a & $ArCH_3$), 1.88 (m, 4H, CH_2CH_2), 1.51 (s, 3H, $ArCH_3$), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.11- 0.10 (m, 9H, 3 x $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 205.7, 176.8, 172.4, 168.4, 164.0, 150.5, 135.2, 111.1, 85.3, 84.8, 84.7, 63.6, 38.0, 37.8, 37.7, 37.3, 29.8, 29.7, 29.0, 27.9, 27.7, 25.9, 18.3, 12.4, -5.42, -5.52. IR (thin film); ν_{max} (cm^{-1}) = 3063, 2954, 2928, 2855, 1739, 1720, 1673, 1537, 1509, 1330, 1253, 1229, 1178, 1033, 835. HRMS (ESI-MS) *m/z* calculated for $C_{62}H_{87}N_4O_{14}Si_2^+$ 1167.5752: found 1167.5740 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(levulinyl)-thymidinyl]}-heptane (**5b**)

Intermediate **3** was prepared according to the procedure for product **5a** as described above. To a solution of **3** (~ 0.48 g, ~0.95 mmol) and **4b** (0.77 g, 1.0 mmol) in MeCN (7.9 mL) was added DBU (0.81 mL, 5.4 mmol) and allowed to stir at 21 ^{o}C. After 20h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL). The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL) and once with brine (50 mL), then dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated in vacuo. The crude product was purified *via* flash column chromatography using MeOH : DCM (1% to 3%) as eluent to afford 0.78 g (68 %) of a colorless foam. R_f (SiO_2 TLC): 0.54 MeOH : DCM (5%). $\lambda_{max(MeCN)}$ = 283 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.92 (d, 1H, H6, *J* = 1 Hz), 7.75 (d, 1H, H6, *J* = 1 Hz), 7.42-7.40 (m, 2H Ar), 7.31-7.21 (m, 7H, Ar), 6.84-6.81 (m, 4H, Ar), 6.41 (dd, 1H, H1'a, *J* = 9.0, 5.5 Hz), 6.33 (dd, 1H, H1'b, *J* = 6 Hz), 5.24 (m, 1H, H3'a), 4.47 (m, 1H, H3'b), 4.38-4.35 (m, 4H, 2 x $ArOCH_2$), 4.13 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.91-3.90 (m, 2H, H5'a & H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 2.74 (m, 2H, CH_2CO), 2.64-2.46 (m, 4H, H2'a, H2'b & $OOCCH_2$), 2.22-2.17 (m, 4H, H2''b & $COCH_3$), 2.00 (m, 1H, H2''a), 1.93 (d, 3H, $ArCH_3$, *J* = 1 Hz), 1.76 (m, 4H, CH_2CH_2), 1.51 (d, 3H, $ArCH_3$, *J* = 1 Hz), 1.38 (m, 6H, $CH_2CH_2CH_2$), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$),0.11 (s, 3H, $SiCH_3$), 0.10 (s, 3H, $SiCH_3$),-0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.5, 170.5, 158.7, 156.0, 155.9, 144.4, 139.4, 139.0, 135.5, 130.1, 130.1, 128.2, 127.90, 127.0, 113.2, 113.2, 104.7, 104.5, 86.6, 86.5, 86.4, 86.2, 85.6, 75.9, 71.1, 67.4, 67.3, 63.6, 62.4, 55.2, 42.2, 39.0, 37.8, 29.8, 29.0, 28.5, 28.5, 28.0, 25.9, 25.9, 25.7, 18.2, 17.9, 12.3, 11.8, -4.64, -4.97, -5.45, -5.54. IR (thin film); ν_{max} (cm^{-1}) = 3063, 2953, 2928, 2855, 2739, 1720, 1673, 1608, 1536, 1509, 1329, 1253, 1178, 1122, 1033, 835. HRMS (ESI-MS) *m/z* calculated for $C_{65}H_{93}N_4O_{14}Si_2^+$ 1209.6221: found 1209.6213 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-thymidinyl]}-butane (**6a**)

A solution (1.5 mL) of 0.5 M $N_2H_4 \bullet H_2O$ in pyridinium acetate (1/1, v/v) was added to **5a** (0.39 g, 0.33 mmol) and allowed to stir. After 5 min, the reaction was diluted with DCM (100 mL) and washed several times with saturated aqueous $NaHCO_3$ (6 x 50 mL), followed by water (50 mL) and brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), concentrated in vacuo and purified *via* flash column chromatography using MeOH : DCM (1% to 5%) as eluent to afford 0.3 g (83 %) of a colorless

foam. R_f (SiO_2 TLC): 0.20 MeOH : DCM (3%). $\lambda_{max(MeCN)}$ =283 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.94 (d, 1H, H6, *J* = 1 Hz), 7.79 (d, 1H, H6, *J* = 1 Hz), 7.41-7.40 (m, 2H Ar), 7.29-7.21 (m, 7H, Ar), 6.83-6.80 (m, 4H, Ar), 6.38 (dd, 1H, H1'a, *J* = 7.5, 5.5 Hz), 6.32 (dd, 1H, H1'b, *J* = 6 Hz), 4.49-4.42 (m, 6H, H3'a, H3'b & 2 x $ArOCH_2$), 4.10 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.91 (m, 1H, H5'a), 3.83 (m, 1H, H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 3.04 (d, 1H, OH, *J* = 3.5), 2.60 (m, 1H, H2'a), 2.47 (m, 1H, H2'b), 2.19 (m, 1H, H2''b), 2.03 (m, 1H, H2''a), 1.93 (d, 3H, $ArCH_3$, *J* = 1 Hz), 1.87 (m, 4H, CH_2CH_2), 1.50 (d, 3H, $ArCH_3$, *J* = 1Hz), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.34, 170.32, 158.7, 156.0, 155.9, 144.4, 139.5, 135.49, 135.48, 130.08, 130.06, 128.2, 127.9, 127.0, 113.2, 113.2, 104.50, 104.45, 87.4, 86.6, 86.5, 86.2, 72.3, 71.0, 66.73, 66.65, 63.5, 62.4, 55.2, 42.23, 42.19, 25.9, 25.7, 25.3, 25.2, 18.3, 17.9, 12.3, 11.7, -4.64, -4.97, -5.42, -5.48. IR (thin film); ν_{max} (cm^{-1}) =3384, 3061, 2954, 2929, 2855, 1669, 1580, 1509, 1472, 1329, 1121, 835. HRMS (ESI-MS) *m/z* calculated for $C_{57}H_{81}N_4O_{12}Si_2^+$ 1069.5384: found 1069.5378 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)- thymidinyl]}-heptane (**6b**)

A solution (2.9 mL) of 0.5 M N_2H_4•H_2O in pyridinium acetate (1/1, v/v) was added to **5b** (1.2 g, 0.98 mmol) and allowed to stir. After 5 min, the reaction was diluted with EtOAc (75 mL) and washed several times with saturated aqueous $NaHCO_3$ (6 x 80 mL), followed by brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), concentrated in vacuo and purified *via* flash column chromatography using MeOH : DCM (1.5% to 3.5%) as eluent to afford 0.74 g (69 %) of a colorless foam. R_f (SiO_2 TLC): 0.10 EtOAc. $\lambda_{max(MeCN)}$ = 283 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.92 (d, 1H, H6, *J* = 1 Hz), 7.78 (d, 1H, H6, *J* = 1 Hz), 7.41-7.40 (m, 2H Ar), 7.29-7.21 (m, 7H, Ar), 6.83-6.81 (m, 4H, Ar), 6.38 (dd, 1H, H1'a, *J* = 7.5, 6.0 Hz), 6.32 (dd, 1H, H1'b, *J* = 6 Hz), 4.48-4.44 (m, 2H, H3'a & H3'b), 4.37-4.34 (m, 4H, 2 x $ArOCH_2$), 4.09 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.92 (m, 1H, H5'a), 3.84 (m, 2H, H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 2.99 (d, 1H, OH, *J* = 3.5 Hz), 2.60 (m, 1H, H2'a), 2.48 (m, 1H, H2'b), 2.20 (m, 1H, H2''b), 2.04 (m, 1H, H2''a), 1.93 (d, 3H, $ArCH_3$, *J* = 1 Hz), 1.73 (m, 4H, CH_2CH_2), 1.51 (d, 3H, $ArCH_3$, *J* = 1 Hz), 1.39 (m, 6H, $CH_2CH_2CH_2$), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$),0.11 (s, 3H, $SiCH_3$), 0.10 (s, 3H, $SiCH_3$),-0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.47, 170.46, 158.7, 156.1, 156.0, 144.4, 139.4, 139.3, 135.5, 130.08, 130.06, 128.2, 127.9, 127.0, 113.20, 113.18, 104.6, 104.5, 87.3, 86.6, 86.6, 86.5, 86.2, 72.3, 71.1, 67.3, 67.3, 63.5, 62.4, 55.2, 42.22, 42.19, 28.9, 28.5, 28.4, 25.9, 25.83, 25.81, 25.7, 18.3, 17.9, 12.3, 11.8, -4.64, -4.97, -5.42, -5.48. IR (thin film); ν_{max} (cm^{-1}) = 3365, 3061, 2953, 2930, 2856, 1671, 1608, 1536, 1509, 1472, 1430, 1328, 1253, 1178, 1121, 835. HRMS (ESI-MS) *m/z* calculated for$C_{60}H_{87}N_4O_{12}Si_2^+$ 1111.5854: found 1111.5820 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-butane (**7a**)

To a solution of **6a** (0.30 g, 0.28 mmol) and DIPEA (92 µL, 0.53 mmol) in THF (2.8 mL) was added Cl-P(OCEt)(N*i*Pr_2) (93 µL, 0.42 mmol) and the reaction was allowed to stir at room temperature. After 20 min, the solvent was removed in vacuo and the crude was taken up in EtOAc (50 mL), washed with aqueous $NaHCO_3$ (3% w/v, 2 x 40mL), followed with brine (40 mL). The organic layer was dried over

anhydrous Na_2SO_4 (~ 1 g), decanted and then evaporated in vacuo. The crude product was purified *via* flash column chromatography using EtOAc (containing 0.1% v/v NEt_3) as eluent to afford 0.27 g (74 %) of a colorless foam. R_f (SiO_2 TLC): 0.53, 0.41 EtOAc. $\lambda_{max(MeCN)}$ = 283 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 7.94 (m, 1H, H6), 7.82-7.79 (m, 1H, H6), 7.51-7.48 (m, 2H Ar), 7.39-7.24 (m, 7H, Ar), 6.92-6.89 (m, 4H, Ar), 6.32-6.25 (m, 2H, H1'a & H1'b), 4.63-4.59 (m, 2H, H3'a & H3'b), 4.41-4.36 (m, 4H, 2 x $ArCH_2$), 4.25, 4.16 (m, 1H, H4'a), 4.03-3.82 (m, 5H, $POCH_2$, H4'b, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.73-3.65 (m, 2H, 2 x NCH), 3.49 (m, 1H, H5'b), 3.35 (m, 1H, H5''b), 2.81-2.77 (m, 2H, CH_2CN), 2.62-2.51 (m, 1H, H2'a), 2.42-2.37 (m, 1H, H2'b), 2.33-2.27 (m, 1H, H2''b), 2.16-2.11 (m, 1H, H2''a), 1.94-1.92 (m, 7H, $ArCH_3$ & CH_2CH_2), 1.62-1.61 (m, 3H, $ArCH_3$), 1.23-1.21 (m, 12H, 4 x CH_3), 0.94-0.93 (m, 9H, $SiC(CH_3)_3$), 0.86-0.85 (m, 9H, $SiC(CH_3)_3$), 0.16-0.14 (m, 6H, 2 x $SiCH_3$), 0.07 (s, 3H, $SiCH_3$), 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.88, 170.87, 159.77, 155.58, 155.57, 155.56, 145.82, 140.87, 140.69, 140.62, 136.54, 136.50, 131.01, 131.00, 129.01, 128.72, 127.77, 118.93, 114.02, 114.01, 104.12, 104.10, 104.01, 87.96, 87.92, 87.57, 87.53, 87.46, 87.34, 87.24, 87.04, 86.87, 75.11, 75.06, 74.98, 74.93, 72.76, 67.10, 64.23, 64.08, 63.80, 59.55, 59.41, 55.57, 44.01, 43.98, 43.91, 43.88, 42.54, 41.49, 41.45, 41.34, 41.31, 26.35, 26.15, 26.04, 24.92, 24.91, 24.89, 24.87, 24.85, 24.83, 24.81, 20.79, 20.77, 20.73, 20.72, 18.93, 18.91, 18.50, 12.48, 12.46, 12.12, -4.44, -4.69, -5.17, -5.18, -5.22.^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.05, 147.68. IR (thin film); ν_{max} (cm^{-1}) = 3064, 2957, 2929, 2856, 1714, 1673, 1608, 1536, 1509, 1471, 1363, 1329, 1253, 1124, 1035, 1002, 835, 782. HRMS (ESI-MS) *m/z* calculated for$C_{66}H_{98}N_6O_{13}PSi_2^+$ 1269.6463: found 1269.6499 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-heptane (**7b**)

To a solution of **6b** (0.31 g, 0.28 mmol) and DIPEA (91 µL, 0.52 mmol) in THF (2.8 mL) was added Cl-P(OCEt)(N*i*Pr_2) (93 µL, 0.42 mmol) and the reaction was allowed to stir at room temperature. After 20 min, the solvent was removed in vacuo and the crude was taken up in EtOAc (50 mL), washed with aqueous $NaHCO_3$ (3% w/v, 2 x 40mL), followed with brine (40 mL). The organic layer dried over anhydrous Na_2SO_4 (~ 1 g), decanted and the solvent evaporated in vacuo. The crude product was purified *via* flash column chromatography using EtOAc (containing 0.1% v/v NEt_3) as eluent to afford 0.32 g (87 %) of a colorless foam. R_f (SiO_2 TLC): 0.81, 0.70 EtOAc. $\lambda_{max(MeCN)}$ = 283 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 7.94-7.93 (m, 1H, H6), 7.81-7.78 (m, 1H, H6), 7.51-7.48 (m, 2H Ar), 7.39-7.24 (m, 7H, Ar), 6.92-6.89 (m, 4H, Ar), 6.32-6.26 (m, 2H, H1'a & H1'b), 4.63-4.60 (m, 2H, H3'a & H3'b), 4.33-4.30 (m, 4H, 2 x $ArCH_2$), 4.25, 4.16 (m, 1H, H4'a), 4.02-3.82 (m, 5H, $POCH_2$, H4'b, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.72-3.65 (m, 2H, 2 x NCH), 3.49 (m, 1H, H5'b), 3.35 (m, 1H, H5''b), 2.80-2.78 (m, 2H, CH_2CN), 2.63-2.51 (m, 1H, H2'a), 2.42-2.37 (m, 1H, H2'b), 2.33-2.28 (m, 1H, H2''b), 2.18-2.11 (m, 1H, H2''a), 1.94-1.93 (m, 3H, $ArCH_3$), 1.79-1.75 (m, 4H, 2 x CH_2), 1.60 (m, 3H, $ArCH_3$), 1.50-1.47 (m, 6H, $CH_2CH_2CH_2$), 1.23-1.21 (m, 12H, 4 x CH_3), 0.94-0.93 (m, 9H, $SiC(CH_3)_3$), 0.86-0.84 (m, 9H, $SiC(CH_3)_3$), 0.16-0.14 (m, 6H, 2 x $SiCH_3$), 0.06 (s, 3H, $SiCH_3$), 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.04, 170.03, 158.87, 154.70, 144.91, 139.92, 139.73, 139.66, 135.64, 135.61, 130.11, 130.10, 128.13, 127.81, 126.87, 118.05, 113.12, 113.10, 103.20, 103.19, 103.09, 87.05, 87.02, 86.66, 86.61, 86.56, 86.41, 86.33, 86.13, 85.93, 74.21, 74.16, 74.08, 74.03, 71.83, 66.53, 66.51, 63.33, 63.18, 62.88, 58.66, 58.51, 54.67, 43.11, 43.08, 43.02, 42.98, 41.63, 40.57, 40.54, 40.42, 40.39, 28.77, 28.32, 25.71, 25.46, 25.26, 24.03, 24.01,

23.99, 23.97, 23.95, 23.94, 23.92, 19.89, 19.88, 19.84, 19.82, 18.03, 18.01, 17.61, 11.60, 11.57, 11.23, -5.33, -5.59, -6.07, -6.08, -6.12. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.04, 147.67. IR (thin film); ν_{max} (cm^{-1}) = 3064, 2957, 2931, 2857, 1714, 1673, 1608, 1536, 1430, 1328, 1254, 1223, 1180, 1126, 1034, 1003, 835, 782. HRMS (ESI-MS) *m/z* calculated for $C_{69}H_{104}N_6O_{13}PSi_2^+$ 1311.6932: found 1311.7026 $[M+H]^+$

The synthesis of phosphoramidites **11a** and **11b** are shown in **Scheme 2**

5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-O^4-hydroxybutyl-thymidine (**8a**)

To a solution of **2** (1.0 g, 2.2 mmol) and triazole (3.4 g, 50 mmol) in MeCN : DCM (14 mL : 14 mL) stirring at 0 °C was added NEt_3 (7.1 mL, 51 mmol) followed by the dropwise addition of $POCl_3$ (0.31 mL, 3.3 mmol). Additional $POCl_3$ (0.20 mL, 2.1 mmol) was added dropwise after 10 min. After 30 min, the solvent was removed in vacuo and the content taken up in DCM (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50mL). The solvent was dried over anhydrous Na_2SO_4, decanted and evaporated to produce the intermediate as a yellow gum. To a solution of this intermediate and 1,4-butanediol (0.22 g, 2.4 mmol) in MeCN (13 mL) was added DBU (0.40 mL, 2.6 mmol). After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50mL) followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and then evaporated to afford a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1 % → 5 %) as eluent to afford 0.50 g (43 % over two steps) of a slightly yellow oil. R_f (SiO_2 TLC): 0.10 EtOAc. $\lambda_{max(MeCN)}$ = 287 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.74 (d, 1H, H6, *J* = 1 Hz), 6.37 (dd, 1H, H1', *J* = 9 Hz & 5 Hz), 5.22 (m, 1H, H3'), 4.40 (t, 2H, $ArOCH_2$, *J* = 6.0 Hz), 4.11 (m, 1H, H4'), 3.89-3.87 (m, 2H, H5' & H5''), 3.68 (m, 2H, $C\underline{H_2}OH$), 2.75 (m, 2H, $COCH_2$), 2.61-2.55 (m, 3H, H2' & $OOCCH_2$), 2.17 (s, 3H, $COCH_3$), 1.99 (m, 1H, H2''), 1.91-1.81 (m, 6H, $ArCH_3$, OH, CH_2), 1.67 (m, 2H, CH_2), 0.87 (s, 9H, $SiC(CH_3)_3$), 0.084 (s, 3H, $SiCH_3$), 0.079 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.4, 155.9, 139.1, 104.7, 86.4, 85.6, 75.9, 67.1, 63.5, 62.2, 39.0, 37.7, 29.8, 29.1, 27.9, 25.8, 25.0, 18.2, 12.2, -5.47, -5.56. IR (thin film); ν_{max} (cm^{-1}) = 3419, 2958, 2929, 2856, 2362, 2335, 1739, 1719, 1669, 1534, 1475, 1432, 1331, 1259, 1231, 1126, 836, 783. HRMS (ESI-MS) *m/z* calculated for $C_{25}H_{43}N_2O_8Si^+$ 527.2783: found 527.2779 $[M+H]^+$.

5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-O^4-hydroxyheptyl-thymidine (**8b**)

To a solution of **2** (2.3 g, 5.1 mmol) and triazole (10.3 g, 148 mmol) in MeCN : DCM (41 mL : 41 mL) stirring at 0 °C was added NEt_3 (21 mL, 151 mmol) followed by the dropwise addition of $POCl_3$ (0.92 mL, 9.9 mmol). Additional $POCl_3$ (0.68 mL, 7.3 mmol) was added dropwise after 15 min. After 30 min, the solvent was removed in vacuo and the content taken up in DCM (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 75mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and then evaporated to produce the intermediate as a yellow gum. To a solution of this intermediate and 1,7-butanediol (0.84 g, 6.3 mmol) in MeCN (37 mL) was added DBU (1.14 mL, 7.6 mmol). After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (70 mL) then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50mL) and followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to afford a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (0.5 % → 2 %) as eluent to

afford 1.2 g (43 % over two steps) of a slightly yellow oil. R_f (SiO_2 TLC): 0.18 EtOAc:Hex (8 : 2). $\lambda_{max(MeCN)}$ = 288 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.72 (d, 1H, H6, *J* = 1 Hz), 6.36 (dd, 1H, H1', *J* = 9 Hz & 5 Hz), 5.21 (m, 1H, H3'), 4.34 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.10 (m, 1H, H4'), 3.90-3.84 (m, 2H, H5' & H5''), 3.60 (m, 2H, $C\underline{H_2}OH$), 2.73 (m, 2H, $COCH_2$), 2.60-2.53 (m, 3H, H2' & $OOCCH_2$), 2.16 (s, 3H, $COCH_3$), 1.97 (m, 1H, H2''), 1.89 (d, 3H, $ArCH_3$, *J* = 1Hz), 1.74-1.68 (m, 3H, OH & CH_2), 1.53 (m, 2H, CH_2), 1.40-1.33 (m, 6H, $CH_2CH_2CH_2$) 0.86 (s, 9H, $SiC(CH_3)_3$), 0.072 (s, 3H, $SiCH_3$), 0.066 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.5, 156.0, 138.9, 104.7, 86.4, 85.6, 75.8, 67.3, 63.5, 62.8, 38.9, 37.7, 32.6, 29.7, 28.9, 28.4, 27.9, 25.8, 25.8, 25.8, 25.6, 18.2, 12.2, -5.49, -5.57. IR (thin film); ν_{max} (cm^{-1}) = 3426, 2961, 2931, 2856, 2362, 2336, 1739, 1721, 1669, 1535, 1473, 1432, 1331, 1259, 1230, 1126, 837, 783. HRMS (ESI-MS) *m/z* calculated for $C_{28}H_{49}N_2O_8Si^+$ 569.3253: found 569.3247 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-thymidinyl]}-butane (**9a**)

To a solution of **8a** (0.50 g, 0.95 mmol), 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine (0.78 g, 0.95 mmol) and Ph_3P (0.50, 1.9 mmol) in dioxane (2.8 mL) was added DIAD (0.37 mL, 1.9 mmol) dropwise over 5 min while stirring. After 16h, additional DIAD (0.1 mL, 0.51 mmol) was added and the reaction was allowed to stir for another 2h. The solvent was removed in vacuo and the content taken up in DCM (50 mL) and washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1 % → 2 %) as eluent to afford 1.2 g of crude material (verified by ^{1}H NMR, data not shown). The product was repurified *via* flash column chromatography using EtOAc : hexanes (3 : 2 → 4: 1) as eluent to afford 0.59 g (46 %) of a colorless foam. R_f (SiO_2 TLC): 0.22 MeOH : DCM (2 %). $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.67 (broad s, 1H, NH), 8.03 (s, 1H, H8), 7.76 (d, 1H, H6, *J* = 1 Hz), 7.39-7.17 (m, 11H, Ar), 7.04-6.99 (m, 3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.42-6.38 (m, 2H, H1'a & H1'b), 5.24 (m, 1H, H3'b), 4.76 (broad s, 2H, $PhOCH_2CO$), 4.65 (m, 2H, $ArOCH_2$), 4.57 (m, 1H, H3'a), 4.45 (m, 2H, $ArOCH_2$), 4.13 (m, 1H, H4'b), 4.06 (m, 1H, H4'a), 3.92-3.86 (m, 2H, H5'b & H5''b), 3.75 (s, 6H, 2 x OCH_3), 3.32 (m, 2H, H5'a & H5''a), 2.78-2.54 (m, 6H, $COCH_2$, $OOCCH_2$, H2'a & H2'b), 2.43 (m, 1H, H2''a), 2.19 (s, 3H, $COCH_3$), 2.02-1.97 (m, 5H, H2''b & CH_2CH_2), 1.92 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$), 0.10 (m, 6H, 2 x $SiCH_3$), 0.03 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.29, 172.40, 170.38, 161.02, 158.48, 157.22, 155.89, 152.39, 151.08, 144.51, 140.12, 139.12, 135.72, 135.68, 129.98, 129.97, 129.77, 128.09, 127.79, 126.83, 122.25, 118.89, 114.91, 113.10, 104.66, 86.95, 86.45, 86.41, 85.64, 84.28, 75.91, 72.50, 68.01, 67.16, 66.84, 63.56, 63.44, 55.17, 40.93, 38.96, 37.75, 29.77, 27.96, 25.85, 25.72, 25.47, 25.21, 18.23, 17.93, 12.27, -4.70, -4.84, -5.45, -5.54. IR (thin film); ν_{max} (cm^{-1}) = 3412, 2956, 2928, 2856, 2361, 2336, 1735, 1719, 1670, 1607, 1539, 1509, 1472, 1437, 1332, 1251, 1177, 1124, 835, 782. HRMS (ESI-MS) *m/z* calculated for $C_{70}H_{92}N_7O_{15}Si_2^+$ 1326.6184: found 1326.6177$[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-thymidinyl]}-heptane (**9b**)

To a solution of **8b** (0.75 g, 1.3 mmol), 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine (0.98 g, 1.2 mmol) and Ph_3P (0.66, 2.5 mmol) in dioxane (4.0 mL) was added DIAD (0.50 mL, 2.5 mmol) dropwise over 5 min while stirring. After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50 mL) and followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and then evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (3 : 2 → 4 : 1) as eluent to afford 0.98 g (71 %) of a colorless foam. R_f (SiO_2 TLC): 0.70 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.64 (broad s, 1H, NH), 8.02 (s, 1H, H8), 7.75 (d, 1H, H6, *J* = 1 Hz), 7.40-7.17 (m, 11H, Ar), 7.05-7.00 (m, 3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.42-6.39 (m, 2H, H1'a & H1'b), 5.24 (m, 1H, H3'b), 4.78 (broad s, 2H, $PhOCH_2CO$), 4.60-4.57 (m, 3H, H3'a & $ArOCH_2$), 4.37 (t, 2H, $ArOCH_2$, *J* = 7 Hz), 4.13 (m, 1H, H4'b), 4.07 (m, 1H, H4'a), 3.93-3.87 (m, 2H, H5'b & H5''b), 3.76 (s, 6H, 2 x OCH_3), 3.33 (m, 2H, H5'a & H5''a), 2.79-2.55 (m, 6H, $COCH_2$, $OOCCH_2$, H2'a & H2'b), 2.44 (H2''a), 2.19 (s, 3H, $COCH_3$), 2.00 (m, 1H, H2''b), 1.92-1.86 (m, 5H, $ArCH_3$ & CH_2), 1.75 (m, 2H, CH_2), 1.52 (m, 2H, CH_2), 1.43-1.42 (m, 4H, CH_2CH_2), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$) 0.104 (s, 3H, $SiCH_3$), 0.098(s, 3H, $SiCH_3$), 0.04 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.30, 172.41, 170.50, 161.21, 158.47, 155.96, 152.33, 151.11, 144.51, 140.02, 138.93, 135.72, 135.69, 129.97, 129.76, 128.09, 127.79, 126.83, 122.23, 118.93, 114.91, 113.10, 113.09, 104.76, 86.94, 86.44, 86.36, 85.60, 84.27, 75.91, 72.52, 68.04, 67.82, 67.43, 63.57, 63.45, 55.17, 40.90, 38.94, 37.76, 29.77, 29.07, 28.78, 28.50, 27.96, 25.90, 25.85, 25.72, 18.23, 17.94, 12.29, -4.70, -4.84, -5.45, -5.54. IR (thin film); ν_{max} (cm^{-1}) = 3412, 2956, 2928, 2856, 2362, 2336, 1735, 1719, 1670, 1607, 1533, 1509, 1464, 1437, 1331, 1250, 1177, 1124, 835, 781. HRMS (ESI-MS) *m/z* calculated for $C_{73}H_{98}N_7O_{15}Si_2^+$ 1368.6654: found 1368.6658 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-thymidinyl]}-butane (**10a**)

A solution (2.5 mL) of 0.5 M N_2H_4•H_2O in pyridinium acetate (1/1, v/v) was added to **9a** (0.75 g, 0.57 mmol). After 15 min of stirring, the reaction was diluted with EtOAc (75 mL added), washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) and then with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1.5 % → 3.5 %) as eluent to afford 0.41 g (58 %) of a colorless foam. R_f (SiO_2 TLC): 0.23 MeOH : DCM (3 %). $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.68 (broad s, 1H, NH), 8.03 (s, 1H, H8), 7.77 (d, 1H, H6, *J* = 1 Hz), 7.40-7.16 (m, 11H, Ar), 7.06-6.99 (m, 3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.42 (dd, 1H, H1'a, *J* = 6.5 Hz), 6.38 (dd, 1H, H1'b, *J* = 7.5 , 6 Hz), 4.76 (broad s, 2H, $PhOCH_2CO$), 4.65 (m, 2H, $ArOCH_2$), 4.57 (m, 1H, H3'a), 4.46-4.43 (m, 3H, H3'b & $ArOCH_2$), 4.08-406 (m, 2H, H4'a & H4'b), 3.91 (dd, 1H, H5'b, *J* = 11, 3 Hz), 3.83 (dd, 1H, H5''b, *J* = 11, 3 Hz), 3.76 (s, 6H, 2 x OCH_3), 3.33-3.32 (m, 2H, H5'a & H5''a), 2.83 (d, 1H, OH, *J* = 4.5 Hz), 2.69 (m, 1H, H2'a), 2.59 (m, 1H, H2'b), 2.44 (m, 1H, H2''a), 2.02-1.97 (m, 5H, H2''b & CH_2CH_2), 1.92 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$), 0.09 (s, 3H, $SiCH_3$), 0.08 (s, 3H, $SiCH_3$), 0.04 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.38, 161.03, 158.48, 157.23, 156.01, 152.39, 151.08, 144.51, 140.11, 139.45, 135.73, 135.69, 129.99, 129.97, 129.77, 128.09, 127.80, 126.83, 122.25, 118.89, 114.92, 113.11, 104.48, 87.31, 86.95, 86.63, 86.45, 84.28, 72.50, 72.39,

68.01, 67.18, 66.80, 63.47, 55.18, 42.22, 40.94, 25.87, 25.72, 25.48, 25.22, 18.30, 17.94, 12.29, -4.69, -4.84, -5.42, -5.49. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3057, 2954, 2929, 2857, 2362, 1724, 1663, 1608, 1532, 1509, 1496, 1462, 1441, 1380, 1362, 1251, 1177, 1067, 940, 835, 782.HRMS (ESI-MS) *m/z* calculated for $C_{65}H_{86}N_7O_{13}Si_2^+$ 1228.5817: found 1228.5822 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-thymidinyl]}-heptane (**10b**)

A solution (3.1 mL) of 0.5 M $N_2H_4 \bullet H_2O$ in pyridinium acetate (1/1, v/v) was added to **9b** (0.95 g, 0.70 mmol). After 20 min of stirring, the reaction was diluted with EtOAc (70 mL added), then washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (2 % → 3.5 %) as eluent to afford 0.64 g (73 %) of a colorless foam. R_f (SiO_2 TLC): 0.68 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.66 (broad s, 1H, NH), 8.02 (s, 1H, H8), 7.78 (d, 1H, H6, *J* = 1 Hz), 7.39-7.17 (m, 11H, Ar), 7.05-6.99 (m, 3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.43-6.37 (m, 2H, H1'a & H1'b), 4.78 (broad s, 2H, $PhOCH_2CO$), 4.58-4.56 (m, 3H, H3'a & $ArOCH_2$), 4.44 (m, 1H, H3'b), 4.36 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.10-406 (m, 2H, H4'a & H4'b), 3.91 (dd, 1H, H5'b, *J* = 11, 3 Hz), 3.82 (dd, 1H, H5''b, *J* = 11, 3 Hz), 3.76 (s, 6H, 2 x OCH_3), 3.33-3.32 (m, 2H, H5'a & H5''a), 3.26 (d, 1H, OH, *J* = 4 Hz), 2.69 (m, 1H, H2'a), 2.61 (m, 1H, H2'b), 2.43 (m, 1H, H2''a), 2.02 (m, 1H, H2''b), 1.94-1.86 (m, 5H, $ArCH_3$ & CH_2), 1.75 (m, 2H, CH_2), 1.52 (m, 2H, CH_2), 1.43-1.42 (m, 4H, 2 x CH_2), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$), 0.09 (s, 3H, $SiCH_3$), 0.08 (s, 3H, $SiCH_3$), 0.04 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 161.2, 158.5, 156.1, 152.3, 151.1, 144.5, 140.0, 139.3, 135.73, 135.70, 130.0, 129.8, 128.1, 127.8, 126.8, 122.2, 118.9, 114.9, 113.1, 113.1, 104.6, 87.4, 86.9, 86.7, 86.4, 84.3, 72.5, 72.3, 68.1, 67.8, 67.4, 63.50, 63.45, 55.2, 42.2, 40.9, 29.1, 28.8, 28.5, 25.9, 25.7, 18.3, 17.9, 12.3, -4.69, -4.84, -5.42, -5.48. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3060, 2953, 2930, 2857, 2361, 2337, 1723, 1663, 1608, 1533, 1508, 1496, 1464, 1437, 1375, 1329, 1250, 1177, 1065, 1034, 940, 834, 781. HRMS (ESI-MS) *m/z* calculated for $C_{68}H_{92}N_7O_{13}Si_2^+$ 1270.6286: found 1270.6294 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-butane (**11a**)

To a solution of compound **10a** (0.20 g, 0.16 mmol) and DIPEA (53 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr_2 (54 µL, 0.25 mmol) dropwise while stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) and followed by brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and then evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.16 g (69 %) of a colorless foam. R_f (SiO_2 TLC): 0.86, 0.81 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 9.33 (s, 1H, NH), 8.22 (s, 1H, H8), 7.82-7.79 (m, 1H, H6), 7.42-7.40 (m, 2H, Ar), 7.35-7.17 (m, 6H, Ar), 7.24-7.16 (m, 3H, Ar), 7.05-6.98 (m, 3H, Ar), 6.91-6.75 (m, 4H, Ar), 6.48-6.45 (m, 1H, H1'a), 6.33-6.29 (m, 1H, H1'b), 5.04 (bs, 2H, $PhOCH_2CO$), 4.90 (m, 1H, H3'a), 4.68-4.60 (m, 3H, H3'b & $ArOCH_2$), 4.45-4.42 (m, 2H,

$ArOCH_2$), 4.26 & 4.16 (m, 1H, H4'b), 4.06-4.04 (m, 1H, H4'a), 3.99-3.83 (m, 4H, H5'b, H5''b & CH_2OP), 3.77-3.66 (m, 8H, 2x (CH)N & 2 x OCH_3), 3.48-3.44 (m, 1H, H5'a), 3.38-3.35 (m, 1H, H5''a), 3.11-3.08 (m, 1H, H2'a), 2.82-2.79 (m, 2H, CH_2CN), 2.64-2.59 (m, 1H, H2'b), 2.50-2.45 (m, 1H, H2''a), 2.19-2.12 (m, 1H, H2''b), 2.08-1.97 (m, 4H, CH_2CH_2, overlapping with acetone solvent residual peak), 1.94 (m, 3H, $ArCH_3$), 1.24-1.21 (m, 12H, 4 x CH_3), 0.94 (m, 9H, $SiC(CH_3)_3$) , 0.87 (m, 9H, $SiC(CH_3)_3$), 0.16-0.15 (m, 6H, 2 x $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.04 (m, 3H, $SiCH_3$) . ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.0, 160.9, 158.6, 158.6, 158.2, 154.8, 154.8, 152.7, 151.5, 145.3, 141.2, 139.8, 139.7, 136.0, 135.9, 130.1, 130.0, 129.5, 128.1, 127.5, 126.5, 121.3, 118.8, 118.1, 114.8, 112.9, 112.8, 103.3, 103.2, 87.1, 87.0, 86.8, 86.7, 86.6, 86.4, 86.2, 86.1, 84.5, 74.21, 74.16, 74.1, 74.0, 72.5, 67.9, 66.7, 66.1, 64.1, 63.3, 63.2, 58.7, 58.5, 54.6, 43.1, 43.1, 43.03, 42.99, 40.6, 40.4, 39.5, 25.5, 25.4, 25.3, 25.1, 24.03, 24.02, 23.98, 23.96, 23.9, 19.90, 19.88, 19.84, 19.83, 18.04, 18.02, 17.6, 11.60, 11.58, -5.36, -5.50, -6.07, -6.12. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.03, 147.68. HRMS (ESI-MS) *m/z* calculated for $C_{74}H_{103}N_9O_{14}PSi_2^+$ 1428.6895: found 1428.6890 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl) -3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-heptane (**11b**)

To a solution of compound **10b** (0.20 g, 0.16 mmol) and DIPEA (52 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr_2 (52 µL, 0.25 mmol) dropwise while stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) followed with brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and then evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.21 g (89 %) of a colorless foam. R_f (SiO_2 TLC): 0.92, 0.90 EtOAc. $\lambda_{max(MeCN)}$ = 272 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 9.30 (s, 1H, NH), 8.23 (s, 1H, H8), 7.80 & 7.80 (m, 1H, H6), 7.43-7.40 (m, 2H, Ar), 7.34-7.16 (m, 6H, Ar), 7.24-7.16 (m, 3H, Ar), 7.05-6.98 (m, 3H, Ar), 6.81-6.75 (m, 4H, Ar), 6.47-6.45 (m, 1H, H1'a), 6.33-6.29 (m, 1H, H1'b), 5.05 (bs, 2H, $PhOCH_2CO$), 4.89 (m, 1H, H3'a), 4.64-4.59 (m, 3H, H3'b & $ArOCH_2$), 4.33-4.31 (m, 2H, $ArOCH_2$), 4.26 & 4.16 (m, 1H, H4'b), 4.07-4.04 (m, 1H, H4'a), 3.99-3.81 (m, 4H, H5'b, H5''b & CH_2OP), 3.75-3.65 (m, 8H, 2x (CH)N & 2 x OCH_3), 3.47-3.44 (m, 1H, H5'a), 3.38-3.35 (m, 1H, H5''a), 3.11-3.07 (m, 1H, H2'a), 2.82-2.78 (m, 2H, CH_2CN), 2.64-2.53 (m, 1H, H2'b), 2.50-2.45 (m, 1H, H2''a), 2.17-2.12 (m, 1H, H2''b), 1.94 (m, 3H, $ArCH_3$), 1.92-1.89 (m, 2H, CH_2), 1.79-1.77 (m, 2H, CH_2), 1.54-1.48 (m, 6H, $(CH_2)_3$), 1.24-1.22 (m, 12H, 4 x CH_3), 0.95-0.94 (m, 9H, $SiC(CH_3)_3$) , 0.87 (m, 9H, $SiC(CH_3)_3$), 0.17-0.15 (m, 6H, 2 x $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.04 (m, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.1, 161.0, 158.6, 158.6, 158.2, 154.8, 152.6, 151.6, 145.3, 141.2, 139.7, 139.7, 136.0, 135.9, 130.1, 130.0, 129.5, 128.1, 127.5, 126.5, 121.3, 118.7, 114.8, 112.9, 112.8, 103.3, 103.2, 86.8, 86.3, 86.14, 86.06, 84.5, 74.0, 72.5, 67.9, 67.1, 66.5, 64.1, 63.3, 63.2, 59.6, 58.7, 58.5, 54.6, 43.1, 43.1, 43.02, 42.99, 39.5, 28.3, 25.7, 25.7, 25.5, 25.3, 24.01, 23.97, 23.95, 19.88, 19.83, 18.0, 17.6, 13.6, 11.6, 11.6, -5.37, -5.51, -6.09, -6.13. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.03, 147.67. HRMS (ESI-MS) *m/z* calculated for $C_{77}H_{109}N_9O_{14}PSi_2^+$ 1470.7365: found 1470.7380 $[M+H]^+$.

The synthesis of phosphoramidites **16a** and **16b** are shown in **Supplementary Scheme 3**

5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-N2-phenoxyacetyl-2'-deoxyguanosine (**13**)

*N*2-Phenoxyacetyl-2'-deoxyguanosine (**12**) (3.0 g, 7.5 mmol) was co-evaporated thrice with anhydrous pyridine (3 x 10 mL). To a solution of the dried residue in pyridine (45 mL) stirring at 0 ^{o}C was added TBS-Cl (1.2 g, 7.8 mmol) which was then allowed to warm slowly to 22 ^{o}C. After 16h, the solvent was removed in vacuo and the content taken up in DCM (100 mL), washed with distilled water (50 mL) followed with a saturated solution of $NaHCO_{3(aq)}$ (75 mL). A gel-like material formed in the organic layer after the second wash, which was then solubilized by the addition of MeOH (5 mL). The solvent was then dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The residue was placed on high vacuum for 2h before being used for the subsequent step. To a solution of the dried residue, EDAC • HCl (2.9 g, 14.9 mmol), and DMAP (cat) in dioxane (54 mL) was added levulinic acid (1.73 g, 14.9 mmol) and allowed to stir at 22 ^{o}C. After 16h, the solvent was removed in vacuo and the content taken up in DCM (50 mL), washed with a saturated solution of $NaHCO_{3(aq)}$ (2 x 50 mL) followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1.0 % → 1.5 %) as eluent to afford 3.1 g (67 % over two steps) of a slightly yellow solid. R_f (SiO_2 TLC):0.46 MeOH : DCM (5 %). $\lambda_{max(MeCN)}$ = 285, 259 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 11.80 (br, 1H, NH), 9.40 (br, 1H, NH), 8.06 (s, 1H, H8), 7.34-7.31 (m, 2H, Ar), 7.05 (m, 1H, Ar), 6.99-6.97 (m, 2H, Ar), 6.26 (dd, 1H, H1', *J* = 8, 6.5 Hz), 5.38 (m, 1H, H3'), 4.70 (s, 2H, $PhOCH_2CO$), 4.16 (m, 1H, H4'), 3.85-3.78 (m, 2H, H5' & H5''), 2.79-2.77 (m, 2H, $COCH_2$), 2.60-2.55 (m, 2H, $OOCCH_2$), 2.10 (s, 3H, $COCH_3$), 0.86 (s, 9H, $SiC(CH_3)_3$), 0.059 (s, 3H, $SiCH_3$), 0.057 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.5, 172.3, 169.7, 156.5, 155.3, 147.6, 146.3, 137.0, 129.9, 122.9, 121.6, 114.9, 85.7, 83.8, 75.5, 67.03, 66.98, 63.6, 38.9, 37.8, 29.8, 27.9, 25.9, 25.8, 18.3, -5.45, -5.61. IR (thin film); ν_{max} (cm^{-1}) = 3417, 2956, 2928, 2856, 2363, 2335, 1717, 1700, 1616, 1559, 1506, 1496, 1473, 1457, 1405, 1360, 1258, 1208, 1155, 1075, 1004, 936, 838, 782. HRMS (ESI-MS) *m/z* calculated for $C_{29}H_{40}N_5O_8Si^+$ 614.2641: found 614.2640 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)- thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (**14a**)

To a solution of **13** (0.76 g, 1.2 mmol), **4a** (1.0 g, 1.4 mmol) and Ph_3P (0.68, 2.6 mmol) in dioxane (4.0 mL) was added DIAD (0.51 mL, 2.6 mmol) dropwise over 5 min while stirring. After 16h, additional DIAD (0.11 mL, 0.54 mmol) was added and the reaction was allowed to stir for another 2h. The solvent was removed in vacuo and the content taken up in EtOAc (75 mL) then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (3 : 2 → 100 % EtOAc) as eluent to afford 0.85 g (52 %) of a colorless foam. R_f (SiO_2 TLC): 0.45 EtOAc. $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.77 (broad s, 1H, NH), 8.23 (s, 1H, H8), 7.93 (d, 1H, H6, *J* = 1 Hz), 7.41-7.21 (m, 11H, Ar), 7.05-7.00 (m, 3H, Ar), 6.84-6.80 (m, 4H, Ar), 6.51 (dd, 1H, H1'a, *J* = 6 & 8 Hz), 6.32 (dd, 1H, H1'b, *J* = 6 Hz), 5.43 (m, 1H, H3'a), 4.77 (broad s, 2H, $PhOCH_2CO$), 4.64 (t, 2H, $ArOCH_2$, *J* = 6 Hz), 4.49-4.43 (m, 3H, $ArOCH_2$ & H3'b), 4.19 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.90-3.88 (m, 2H, H5'a & H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (dd, 1H, H5'b, J= 3 & 11 Hz), 3.25 (d, 1H, H5''b, *J* = 3 & 11 Hz), 2.78 (m, 2H, $OOCCH_2$), 2.67-2.61 (m, 4H, CH_2CO, H2'a & H2''a), 2.48 (m, 1H, H2'b), 2.22-2.17 (m, 4H, H2''b & $COCH_3$), 2.04-1.96 (m, 4H, CH_2CH_2),

1.50 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.79 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.3, 161.0, 158.7, 157.2, 155.9, 152.5, 151.2, 144.4, 139.8, 139.5, 135.49, 135.47, 130.08, 130.06, 129.8, 128.2, 127.9, 127.0, 122.3, 118.6, 114.9, 113.2, 113.2, 104.5, 86.6, 86.5, 86.2, 85.7, 84.0, 75.7, 71.0, 68.0, 67.2, 66.7, 63.7, 62.4, 55.2, 42.2, 39.1, 37.8, 29.8, 28.0, 26.0, 25.7, 25.4, 25.2, 18.4, 17.9, 11.7, -4.64, -4.97, -5.40, -5.56. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3060, 2956, 2929, 2856, 2362, 2336, 1734, 1718, 1669, 1607, 1532, 1509, 1472, 1437, 1329, 1301, 1251, 1177, 1154, 1070, 836, 781. HRMS (ESI-MS) *m/z* calculated for $C_{70}H_{92}N_7O_{15}Si_2^+$ 1326.6184: found 1326.6191 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (**14b**)

To a solution of **13** (0.71 g, 1.2 mmol), **4b** (1.0 g, 1.3 mmol) and Ph_3P (0.64, 2.5 mmol) in dioxane (3.8 mL) was added DIAD (0.48 mL, 2.5 mmol) dropwise over 5 min while stirring. After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (75 mL) then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (1 : 1 → 4 : 1) as eluent to afford 0.76 g (48 %) of a colorless foam. R_f (SiO_2 TLC): 0.70 EtOAc. $\lambda_{max(MeCN)}$ = 270 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 8.75 (broad s, 1H, NH), 8.22 (s, 1H, H8), 7.91 (d, 1H, H6, *J* = 1 Hz), 7.68-7.20 (m, 11H, Ar), 7.04-7.01 (m, 3H, Ar), 6.84-6.80 (m, 4H, Ar), 6.51 (dd, 1H, H1'a, *J* = 6 & 9 Hz), 6.32 (dd, 1H, H1'b, *J* = 6 Hz), 5.42 (m, 1H, H3'a), 4.79 (broad s, 2H, $PhOCH_2CO$), 4.58 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.47 (m, 1H, H3'b), 4.35 (t, 2H, $ArOCH_2$, *J* = 7 Hz), 4.19 (m, 1H, H4'a), 3.95 (m, 1H, H4'b), 3.90-3.88 (m, 2H, H5'a & H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (dd, 1H, H5'b, J= 3 & 11 Hz), 3.24 (dd, 1H, H5''b, *J* = 3 & 11 Hz), 2.78 (m, 2H, $OOCCH_2$), 2.67-2.61 (m, 4H, CH_2CO, H2'a & H2''a), 2.48 (m, 1H, H2'b), 2.22-2.17 (m, 4H, H2''b & $COCH_3$), 1.89 (m, 2H, CH_2), 1.75-1.71 (m, 4H, 2 x CH_2), 1.50 (d, 3H, $ArCH_3$, *J* = 1 Hz), 1.43-1.40 (m, 4H, 2 x CH_2), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.79 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.02 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.5, 161.2, 158.6, 155.9, 152.5, 151.2, 144.4, 139.7, 139.3, 135.5, 132.1, 132.0, 130.08, 130.06, 129.8, 128.5, 128.4, 128.2, 127.9, 127.0, 122.3, 118.6, 114.9, 113.2, 113.2, 104.6, 86.6, 86.5, 86.2, 85.7, 84.0, 75.7, 71.1, 68.0, 67.8, 67.3, 63.7, 62.4, 55.2, 42.2, 39.1, 37.8, 29.8, 29.1, 28.9, 28.5, 28.0, 26.0, 25.9, 25.7, 18.4, 17.9, 14.2, 11.8, -4.64, -4.97, -5.40, -5.56. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3059, 2954, 2930, 2856, 2362, 2336, 1734, 1718, 1669, 1607, 1529, 1509, 1472, 1437, 1378, 1328, 1301, 1252, 1177, 1155, 1071, 836, 782, 734. HRMS (ESI-MS) *m/z* calculated for $C_{73}H_{98}N_7O_{15}Si_2^+$ 1368.6654: found 1368.6661 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (**15a**)

A solution (2.8 mL) of 0.5 M $N_2H_4 \bullet H_2O$ in pyridinium acetate (1/1, v/v) was added to **14a** (0.85 g, 0.64 mmol). After 30 min of stirring, the reaction was diluted with EtOAc (70 mL added), washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1.5 % → 2.5 %) as eluent to

afford 0.57 g (73 %) of a colorless foam. R_f (SiO_2 TLC):0.17 MeOH : DCM (4 %). $\lambda_{max(MeCN)}$ = 270 nm.. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.85 (broad s, 1H, NH), 8.23 (s, 1H, H8), 7.95 (d, 1H, H6, *J* = 1 Hz), 7.42-7.40 (m, 2H, Ar), 7.36-7.23 (m, 9H, Ar), 7.07-7.01 (m, 3H, Ar), 6.84-6.82 (m, 4H, Ar), 6.70 (dd, 1H, H1'a, *J* = 6.8 Hz), 6.33 (dd, 1H, H1'b, *J* = 6 Hz), 4.77 (m, 3H, H3'a & $PhOCH_2CO$), 4.62 (t, 2H, $ArOCH_2$, *J* = 6 Hz), 4.50-4.43 (m, 3H, $ArOCH_2$ & H3'b), 4.17 (m, 1H, H4'a), 3.97 (m, 1H, H4'b), 3.89-3.88 (m, 2H, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.52 (dd, 1H, H5'b, J= 3 & 11 Hz), 3.26 (dd, 1H, H5''b, *J* = 3 & 11 Hz), 2.64-2.61 (m, 2H, H2'a & H2''a), 2.49 (m, 1H, H2'b), 2.20 (m, 1H, H2''b), 2.03-1.96 (m, 4H, CH_2CH_2), 1.51 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.91 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), 0.00 (s, 3H, $SiCH_3$), - 0.07 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 161.0, 158.7, 157.1, 155.9, 152.4, 150.9, 144.4, 140.4, 139.5, 135.49, 135.47, 130.09, 130.06, 129.8, 128.2, 127.9, 127.1, 122.4, 118.8, 114.9, 113.21, 113.18, 104.6, 87.8, 86.6, 86.5, 86.2, 84.3, 72.5, 71.0, 67.9, 67.1, 66.7, 63.9, 62.3, 58.5, 55.2, 42.2, 41.6, 29.7, 26.0, 25.7, 25.4, 25.2, 18.4, 17.9, 11.7, 8.26, -4.63, -4.96, -5.36, -5.50. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3059, 2957, 2928, 2856, 2362, 2335, 1717, 1663, 1608, 1532, 1509, 1472, 1437, 1381, 1329, 1301, 1251, 1176, 1150, 1064, 1034, 835, 781, 734. HRMS (ESI-MS) *m/z* calculated for $C_{65}H_{86}N_7O_{13}Si_2^+$ 1228.5817: found 1228.5821 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (**15b**)

A solution (2.4 mL) of 0.5 M $N_2H_4 \bullet H_2O$ in pyridinium acetate (1/1, v/v) was added to **14b** (0.74 g, 0.54 mmol). After 30 min of stirring, the reaction was diluted with EtOAc (70 mL added), then washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (~ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1.5 % → 2.5 %) as eluent to afford 0.55 g (80 %) of a colorless foam. R_f (SiO_2 TLC):0.23 MeOH : DCM (4 %). $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.80 (broad s, 1H, NH), 8.21 (s, 1H, H8), 7.92 (d, 1H, H6, *J* = 1 Hz), 7.42-7.40 (m, 2H, Ar), 7.36-7.21 (m, 9H, Ar), 7.06-7.01 (m, 3H, Ar), 6.84-6.81 (m, 4H, Ar), 6.66 (dd, 1H, H1'a, *J* = 6.5 Hz), 6.33 (dd, 1H, H1'b, *J* = 6 Hz), 4.75 (m, 3H, H3'a & $PhOCH_2CO$), 4.57 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.47 (m, 1H, H3'b), 4.35 (t, 2H, $ArOCH_2$, *J* = 7 Hz), 4.15 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.89-3.88 (m, 2H, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.52 (dd, 1H, H5'b, J= 3 & 11.5 Hz), 3.26 (dd, 1H, H5''b, *J* = 3 & 11.5 Hz), 2.64-2.61 (m, 2H, H2'a & H2''a), 2.49 (m, 1H, H2'b), 2.20 (m, 1H, H2''b), 1.88 (m, 2H, CH_2), 1.74 (m, 2H, CH_2), 1.51 (m, 5H, CH_2 & $ArCH_3$), 1.43-1.41 (m, 4H, CH_2CH_2) 0.91 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.01 (s, 3H, $SiCH_3$), - 0.07 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 161.2, 158.7, 156.0, 152.3, 151.0, 144.4, 140.3, 139.3, 135.5, 132.1, 132.0, 130.09, 130.07, 129.8, 128.5, 128.4, 128.2, 127.9, 127.0, 122.4, 118.8, 114.9, 113.20, 113.18, 104.6, 87.6, 86.6, 86.5, 86.2, 84.2, 72.5, 71.1, 68.0, 67.8, 67.3, 67.1, 63.8, 62.4, 55.2, 42.2, 41.5, 29.1, 28.8, 28.5, 26.0, 25.9, 25.7, 18.4, 17.9, 11.8, -4.64, -4.96, -5.36, -5.50. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3059, 2953, 2930, 2856, 2362, 2336, 1717, 1668, 1607, 1531, 1509, 1471, 1437, 1379, 1328, 1302, 1251, 1176, 1110, 1066, 1034, 835, 781, 733. HRMS (ESI-MS) *m/z* calculated for $C_{68}H_{91}N_7O_{13}Si_2Na^+$ 1292.6106: found 1292.6110 $[M+Na]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (**16a**)

To a solution of compound **15a** (0.20 g, 0.16 mmol) and DIPEA (53 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr$_2$ (54 µL, 0.25 mmol) dropwise while stirring. After 35 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with 3% (w/v) $NaHCO_{3(aq)}$ (2 x 35 mL) followed by brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using an EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.19 g (83 %) of a colorless foam. R_f (SiO_2 TLC): 0.84, 0.76 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm)): 9.43 (s, 1H, NH), 8.28-8.27 (s, 1H, H8), 7.94 (m, 1H, H6), 7.51-7.49 (m, 2H, Ar), 7.40-7.25 (m, 9H, Ar), 7.06-6.97 (m, 3H, Ar), 6.93-6.90 (m, 4H, Ar), 6.51-6.47 (m, 1H, H1'a), 6.29-6.27 (m, 1H, H1'b), 5.10 (s, 2H, $PhOCH_2CO$), 4.89-4.83 (m, 1H, H3'a), 4.69-4.61 (m, 3H, H3'b & $ArOCH_2$), 4.44-4.41 (m, 2H, $ArOCH_2$), 4.26 & 4.17 (m, 1H, H4'a), 4.03-3.62 (m, 13H, H4'b, H5'a, H5''a, CH_2OP, 2x (CH)N & 2 x OCH_3), 3.51-3.48 (m, 1H, H5'b), 3.37-3.34 (m, 1H, H5''b), 3.07-2.98 (m, 1H, H2'a), 2.82-2.61 (m, 3H, H2''a & CH_2CN), 2.43-2.38 (m, 1H, H2'b), 2.34-2.29 (m, 1H, H2''b) , 2.05-2.02 (m, 2H, CH_2, overlapping with acetone solvent residual peak), 2.00-1.96 (m, 2H, CH_2), 1.61 (m, 3H, $ArCH_3$), 1.25-1.20 (m, 12H, 4 x CH_3), 0.93-0.91 (m, 9H, $SiC(CH_3)_3$) , 0.87-0.85 (m, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.07 (m, 3H, $SiCH_3$), 0.02 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.0, 166.92, 160.90, 158.9, 158.2, 154.7, 152.80, 151.79, 144.9, 140.6, 140.5, 140.0, 135.7, 135.6, 130.1, 129.4, 128.1, 127.8, 126.9, 121.2, 118.4, 118.1, 114.7, 113.1, 103.3, 87.1, 86.9, 86.8, 86.6, 86.4, 86.0, 84.4, 74.1, 73.9, 73.8, 73.7, 71.8, 67.9, 67.4, 66.7, 66.1, 65.7, 63.4, 63.2, 62.9, 61.3, 59.6, 58.8, 58.7, 58.5, 54.7, 43.2, 43.12, 43.06, 43.02, 41.6, 39.1, 38.8, 25.5, 25.4, 25.3, 25.1, 24.1, 24.03, 24.00, 23.98, 19.92, 19.87, 18.1, 17.6, 13.6, 11.2, -5.33, -5.59, -6.06, -6.09. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.24, 148.13. HRMS (ESI-MS) *m/z* calculated for $C_{74}H_{103}N_9O_{14}PSi_2^+$ 1428.6895: found 1428.6898 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-N2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (**16b**)

To a solution of compound **15b** (0.20 g, 0.16 mmol) and DIPEA (52 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr$_2$ (52 µL, 0.25 mmol) dropwise while stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) followed with brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and then evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using an EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.19 g (82 %) of a colorless foam. R_f (SiO_2 TLC): 0.87, 0.86 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm)): 9.42 (s, 1H, NH), 8.28-8.27 (s, 1H, H8), 7.94 (m, 1H, H6), 7.51-7.50 (m, 2H, Ar), 7.36-7.25 (m, 9H, Ar), 7.05-7.03 (m, 2H, Ar), 7.00-6.97 (m, 1H, Ar), 6.92-6.91 (m, 4H, Ar), 6.50-6.47 (m, 1H, H1'a), 6.30-6.27 (m, 1H, H1'b), 5.11 (s, 2H, $PhOCH_2CO$), 4.88-4.84 (m, 1H, H3'a), 4.64-4.59 (m, 3H, H3'b & $ArOCH_2$), 4.33-4.30 (m, 2H, $ArOCH_2$), 4.24 & 4.17 (m, 1H, H4'a), 4.09-3.80 (m, 11H, H4'b, H5'a, H5''a, CH_2OP & 2 x OCH_3), 3.72-3.65 (m, 2H, 2x (CH)N), 3.51-3.48 (m, 1H, H5'b), 3.37-3.34 (m, 1H, H5''b), 3.07-2.98 (m, 1H, H2'a), 2.82-

2.76 (m, 2H, CH_2CN), 2.74-2.61 (m, 1H, H2''a), 2.43-2.38 (m, 1H, H2'b), 2.34-2.29 (m, 1H, H2''b) , 1.92-1.87 (m, 2H, CH_2) 1.78-1.74 (m, 2H, CH_2), 1.61 (m, 3H, $ArCH_3$), 1.55-1.48 (m, 6H, $(CH_2)_3$), 1.26-1.21 (m, 12H, 4 x CH_3), 0.92-0.91 (m, 9H, $SiC(CH_3)_3$) , 0.86-0.85 (m, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.07 (m, 3H, $SiCH_3$), 0.02 (s, 3H, $SiCH_3$).^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.1, 161.0, 158.9, 158.3, 154.8, 152.8, 151.8, 144.9, 140.5, 140.4, 139.9, 135.7, 135.6, 130.12, 130.11, 129.4, 128.1, 127.8, 126.9, 121.2, 118.4, 118.1, 114.7, 113.13, 113.11, 103.3, 87.1, 86.8, 86.6, 86.4, 86.0, 84.4, 84.3, 74.1, 73.9, 73.8, 73.7, 71.8, 67.9, 67.1, 66.5, 63.4, 63.2, 62.9, 59.6, 58.8, 58.7, 58.5, 54.7, 43.2, 43.12, 43.06, 43.0, 41.6, 39.1, 38.8, 28.3, 25.72, 25.71, 25.49, 25.48, 25.3, 24.1, 24.04, 24.01, 23.98, 23.95, 19.93, 19.90, 19.87, 19.85, 18.1, 17.6, 13.6, 11.2, -5.33, -5.58, -6.04, -6.06, -6.08, -6.10. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.23, 148.13. HRMS (ESI-MS) *m/z* calculated for $C_{77}H_{109}N_9O_{14}PSi_2^+$ 1470.7365: found 1470.7375 $[M+H]^+$.

Preparation, purification and characterization of the IaCL oligonucleotide

The cross-linked duplexes, the sequences for which are shown in **Figure 1** of the main text, were assembled with an Applied Biosystems Model 3400 synthesizer on a 1.5 mmol scale using β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites protected with "fast-deprotecting" groups were dissolved in anhydrous MeCN at a concentration of 0.1 M for the 3'-*O*-deoxyphosphoramidites, and 0.13-0.15 M for the cross-linked 3'-*O*-deoxyphosphoramidites. Oligomer sequence assembly was carried out as previously described.[6] The capping step of the assembly was carried out using phenoxyacetic anhydride/pyridine/tetrahydrofuran 1:1:8 (v/v/v; solution A) and 1-methyl-imidazole/ tetrahydrofuran 16:84 (w/v; solution B). Coupling wait times for phosphoramidites **7a**, **7b**, **11a**, **11b**, **16a**, and **16b** were extended to 10 min (compared to 2 min for the commercially available phosphoramidites). Protecting group removal and cleavage from the solid support was carried out by treatment with 0.05M K_2CO_3 in MeOH for 4 h at room temperature with mild rocking in 2 mL screw-cap microfuge tubes fitted with Teflon lined caps. The base was neutralized with an equimolar amount of AcOH and crude oligomers were transferred and lyophilized in a Speedvac concentrator. Silyl protecting groups were removed from oligomers by treatment with NEt_3•3HF (200 µL, pellet initially sonicated (2 x 15 s)) for 16h at room temperature under gentle rocking. Oligomers were precipitated using cool *n*-butanol (400 µL) and the resulting mixture was allowed to cool to -20^{o}C for 10 min, followed by spinning the samples down. The supernatant was removed and the pellet was washed with another aliquot of *n*-butanol (400 µL). Purification was achieved by strong anion exchange HPLC using a Dionex DNAPAC PA-100 column (0.4 cm x 25 cm) purchased from Dionex Corp, Sunnyvale, CA using a linear gradient of 0–52% buffer B over 24 min (buffer A: 100 mm Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mm Tris HCl, pH 7.5, 10% MeCN, 1M NaCl) at 55 ^{o}C. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters Inc.) as previously described.[9] The molecular mass of the modified oligomers were identified by ESI-MS and the measured values were in agreement with the expected masses (**Supplementary Information Figure S52 - S57** for MS spectra).

UV thermal denaturation

Molar extinction coefficients for the unmodified and cross-linked oligonucleotides were calculated from those of the mononucleotides and dinucleotides using the nearest-neighbor approximations (M^{-1} cm^{-1}). All duplexes were prepared by mixing equimolar amounts of the interacting strands and lyophilizing the mixture to dryness. The resulting pellet was then dissolved in 90mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA buffer (pH 7.0) to give a final concentration of 3.5 µM duplex. Prior to the thermal run, samples were degassed by placing them in a speed- vac concentrator for 2 min. Annealing curves were acquired at 260 nm starting at 95 ^{o}C and decreasing temperature at a rate of cooling of 0.5 ^{o}C min^{-1} until 15 ^{o}C, using a Varian CARY Model 3E spectrophotometer fitted with a 6-sample thermostated cell block and a temperature controller. The samples were then denatured by heating from 15 to 95 ^{o}C at an increasing temperature rate of 0.5 ^{o}C min^{-1} to show reversibility. Denaturing data processing was carried out as described by Puglisi and Tinoco[10] and transferred to Microsoft Excel™ for viewing.

Circular dichroism (CD) spectroscopy

Circular dichroism spectra were obtained on a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath. Samples were allowed to equilibrate for 5 – 10 min at 15 ^{o}C in 90 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA (pH 7.0), at a final concentration of 3.4 µM. Each spectrum was an average of 3 scans, collecting at a rate of 50 nm min^{-1} , with a bandwidth of 1 nm and sampling wavelength of 0.2 nm using fused quartz cells (Starna 29-Q-10). The CD spectra were recorded from 350 to 220 nm at 15 ^{o}C. The molar ellipticity (ϕ) was calculated from the equation $\phi = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the molar concentration of oligonucleotides (moles/L), and l is the path length of the cell (cm). The data were processed using software supplied by the manufacturer (JASCO, Inc.) and transferred into Microsoft Excel™ for viewing.

Molecular modelling

Molecular modeling was performed by using the Hyperchem 7.5 software package from Hypercube utilizing the AMBER force field. Hybridized oligomers containing a dGG·dCC, dGpG4·dCC, dGpG4·dCC base pair were constructed from the nucleic acid template option using a B-form duplex. Sequence contexts were truncated to the following 5'-GGC TXX GAT C and 3'-CCG AYY CTA G for proper solvation. Duplexes were solvated with water using a periodic box occupying at least 3 times the volume of the duplex alone. Standard Amber99 parameters were used with the dielectric set to constant. "One to four scale factors" non-bonded interactions were set to 0.5 (both electrostatic and van der Waals). Cutoffs were applied to "switched" to an outer and inner radius of 14.5 and 10.5 Å. All structures were geometry optimized using Polak-Ribiere conjugate gradient until the RMS gradient was less than 0.1 kcal/(Å mol) using the periodic boundary condition option.

Protein expression and purification

Ampicillin, isopropyl β-D-thiogalactopyranoside (IPTG), and most other biochemical reagents as well as polyacrylamide gel materials were purchased from Bioshop Canada Inc (Burlington, ON). Ni-NTA Superflow Resin was purchased from Qiagen (Mississauga, ON). Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC). Nitro-cellulose filters (0.20 μm) were obtained from Millipore. XL-10 Gold and Bl21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). T4 polynucleotide kinase (PNK) was obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). Phusion Polymerase was obtained from New England Biolabs (Ipswich, MA). DNA primers for site directed mutagenesis and cloning were purchased from Biocorp (Montreal, QC). All AGT homologues were expressed under the promoter of the pQE30 vector in XL-10 Gold *E. coli* cells, as previously described.[2,4,5,11]

AGT repair assay of IaCL DNA duplexes

DNA substrates were labeled at the using γ-[^{32}P]-ATP as previously described.[4] Briefly, a 20 μM solution of DNA was made in 1X PNK buffer along with 1 μL γ-[^{32}P]-ATP (10 μCi μL^{-1}) and 5 units of T4 PNK. The labelling reaction was conducted for 1 h at 37 °C after which the reaction was terminated by boiling the sample for 5 min. 100 pmol of labeled DNA was added to 110 pmol of the complement strand in a total volume of 50 μL of water making a 2 μM dsDNA solution with 10% excess of the non-damaged strand. The solution was boiled for 2 min, cooled slowly to room temperature and kept in a refrigerator at 4 °C overnight to ensure proper annealing of the duplex. The repair reaction mixtures were constituted of 2 pmol of the DNA duplex and 10 pmol or 60 pmol of AGT in a total volume of 15 μL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] and allowed to react at 37 °C overnight. The reaction was terminated by the addition of 18.2 μL of stop buffer [81 mM Tris–HCl, 81 mM boric acid, 1.8 mM EDTA and 1% SDS (sodium dodecyl sulfate) (pH 8.0) in 80% formamide] followed by boiling for 2 min. Samples (10 μL) were loaded on a 14 cm × 16 cm, 20% 7 M urea denaturing polyacrylamide gel (19 : 1) for separation. The gels were run using 1X TBE for 1h-1.5h at 400 V and the gels exposed to a storage phosphor screen. The image was captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and the autoradiography counts obtained by Image-Quant™ (Amersham Biosciences). For the repair time course assays, master mixes of 150 μL composed of 100 pmol (5-fold) or 600 pmol (30-fold) of AGT and 20 pmol of DNA substrate were prepared. Each sample was placed at 37 ^{o}C and at each time point, 7.5 μL was removed from the master mix and placed in a tube with 9.1 μL of stop reaction buffer and analysis was achieved as described above. All samples were analyzed on polyacrylamide gel electrophoresis (using 1X TBE for 1h-1.5h at 400 V) and visualized as described above.

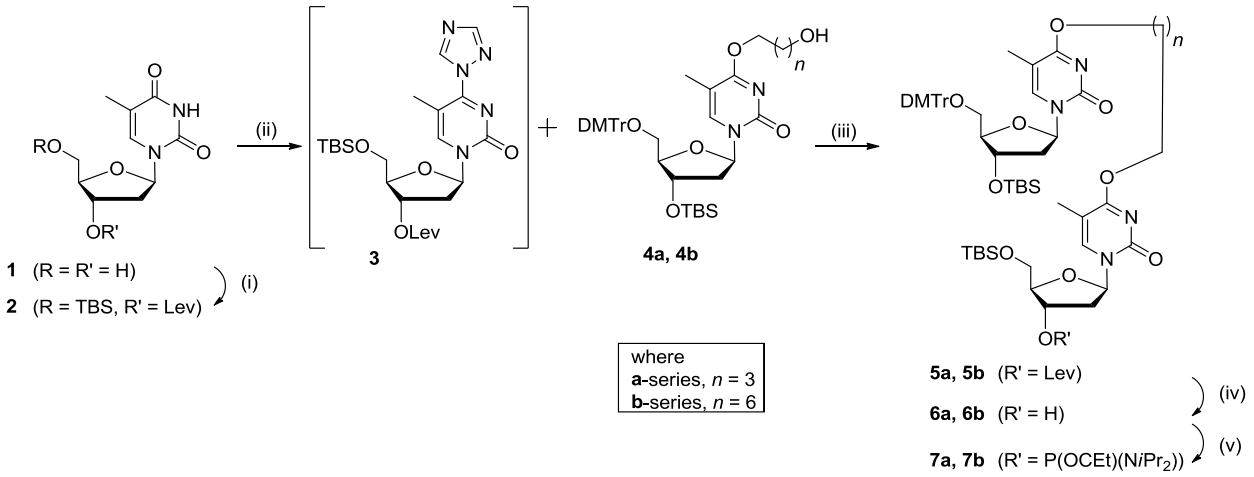
Identification of IaCL Repair Product by SDS-PAGE

600 pmol of hAGT or Ada-C was incubated with 600 pmol of IaCL (**GG4**, **GG7**, **TG4** or **TG7** in 25 μL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] overnight at 37 °C. The reaction was terminated by adding 6 μL of SDS Buffer [200 mM Tris-HCl (pH 7.6), 3.2% SDS, 48% glycerol and 0.4 M DTT], boiling the sample for 2 min, and separated on a 15% SDS- PAGE. The gel was stained with Coomassie Blue Stain as per the manufacturer's protocol. Detection and visualization of DPC products

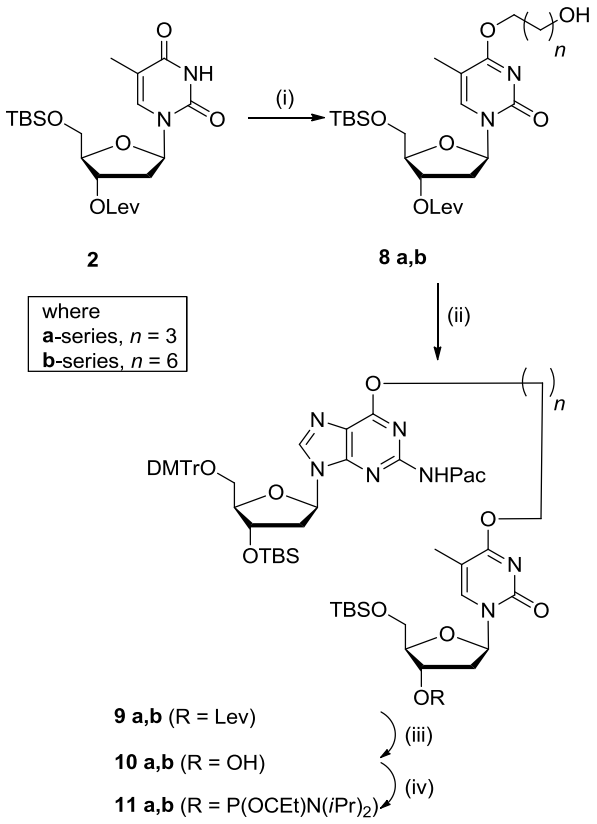
was conducted with Typhoon Trio Variable Mode Imager in fluorescence mode with the green (532 nm) laser in conjunction with the 526-nm short-pass filter.

Identification of repair product (DPC) from hAGT-mediated repair of IaCL DNA

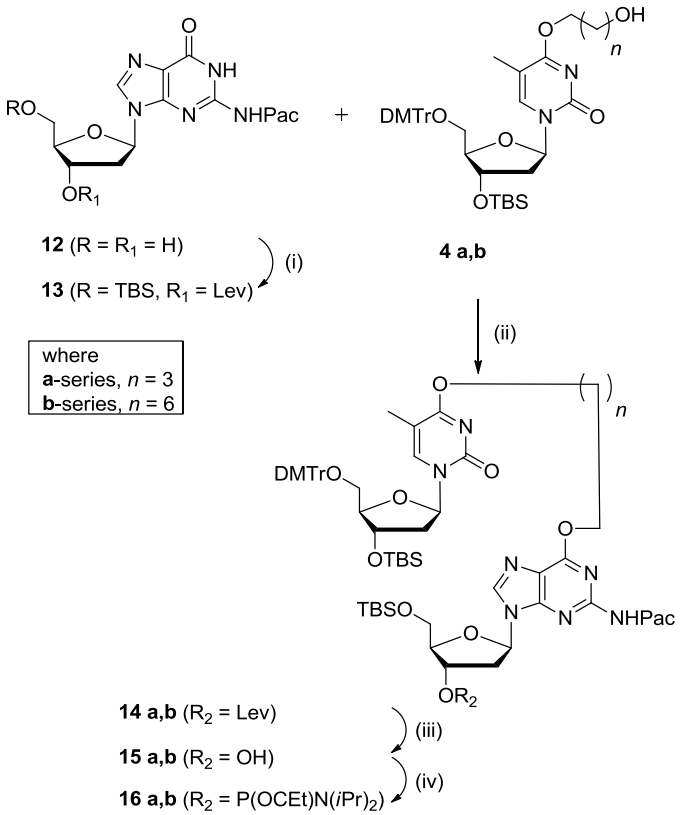
600 pmol of hAGT was incubated with 600 pmol of **GG4**, **GG7**, **TG4** or **TG7** in 25 µL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] for 30 min at 37 °C. A reaction aliquot (5 µL) was diluted in aqueous 0.1 % (v / v) formic acid subject to HPLC (Agilent 1200 Series system) using a Grace™ Vydac™ C4 (214MS) column (Fisher Scientific) (100mm x 2.1 mm) operated at a flow rate of 0.25 mL/min at room temperature (22 ^{o}C) with the following gradient method: 0 - 0.1 min, linear gradient from 10 - 20 % B, 0.1 - 6 min, linear gradient from 20 - 95 % B, 6 - 8 min, hold at 95 % B, 8 - 10 min, linear gradient from 95 – 10 % B, 10 - 19 min, hold at 10 % B (buffer A, 0.1% formic acid in water and buffer B, 0.1% formic acid in acetonitrile). The LC system was interfaced to a Micromass Q-Tof Ultima API equipped with an electrospray source set with the following conditions: source voltage 3.5 kV, mass range of 700−1999 m/z in positive ion mode. The theoretical mass of the hAGT-DNA species were calculated by summing the mass of hAGT (21876 Da), either a butylene linker (54 Da) or heptylene linker (96 Da), and the corresponding fragments GGCTT (1494 Da) or GGCTG (1519 Da). The expected masses are reported in **Figure 7** of the main text. If the repair event had occurred at 5'-residue of the IaCL adduct, DPC masses would have been significantly different since the DNA fragment differed in size and sequence context (GGA TCA CCAG (3037 Da)).



Supplementary Scheme 1: Synthesis of O^4-dT-alkylene-O^4-dT dimer phosphoramidites **7a** and **7b**. Reagents and conditions where **a**-compounds are *n* = 3 and **b**-compounds are *n* = 6: (i) TBS-Cl (1.1 eq), Py, 0°C, 4h, work up, Levulinic acid (2 eq), EDC (2 eq), dioxane, 21°C, 16h. (ii) Triazole (22.5 eq), NEt_3 (23eq), $POCl_3$ (2.1 eq), MeCN:DCM (1:1 v / v), 0°C, 2h. (iii) DBU (2.3 eq), MeCN, 21°C, 16h. (iv) $N_2H_4 \bullet H_2O$ (0.5 M) in pyridinium acetate buffer (1:1 v / v), 21°C, 10 min. (v) Cl-P(OCEt)(N*i*Pr$_2$) (1.5 eq), DIPEA (1.9 eq), THF, 21°C, 30 min. It should be noted that the synthesis of compounds **4a** and **4b** has been reported previously.[2]

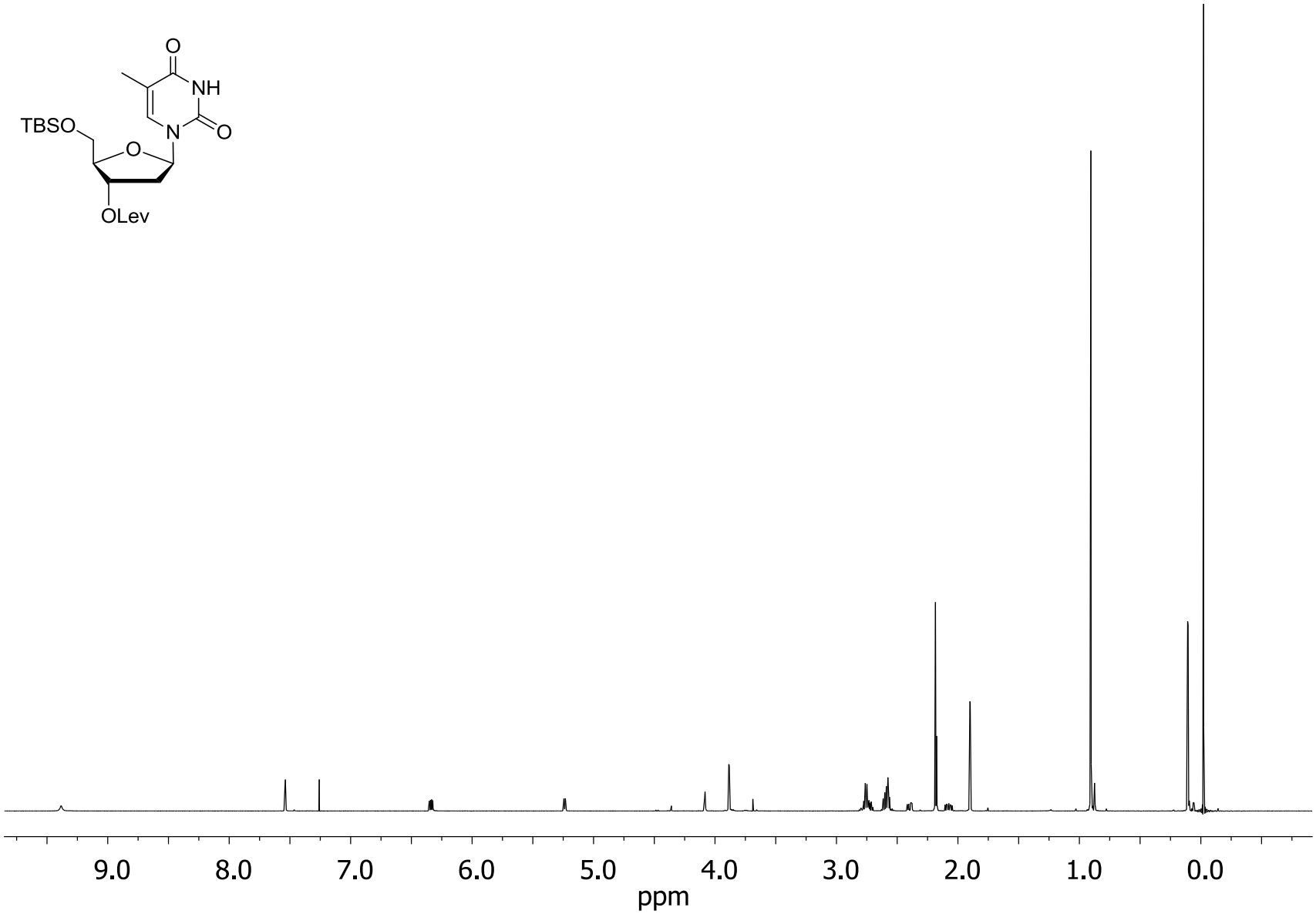


Supplementary Scheme 2: Synthesis of O^6-dG-alkylene-O^4-dT dimer phosphoramidites **11a** and **11b**. Reagents and conditions: (i) 1. Triazole (22.5 eq) NEt_3 (23 eq), $POCl_3$ (2.5 eq), DCM : MeCN (1:1 v / v), 0°C, 30 min, 2. 1,4-butanediol or 1,7-heptanediol (1.1 eq), DBU (1.2 eq), MeCN, 21°C, 16h. (ii) 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine (1.1 eq), Ph_3P (2 eq), DIAD (2.5 eq), 21 °C, 18h. (iii) N_2H_4 • H_2O (0.5 M) in pyridinium acetate buffer (1:1 v / v), 21°C, 15 min. (iv) Cl-P(OCEt)(N*i*Pr_2) (1.5 eq), DIPEA (1.9 eq), THF, 21°C, 30 min.

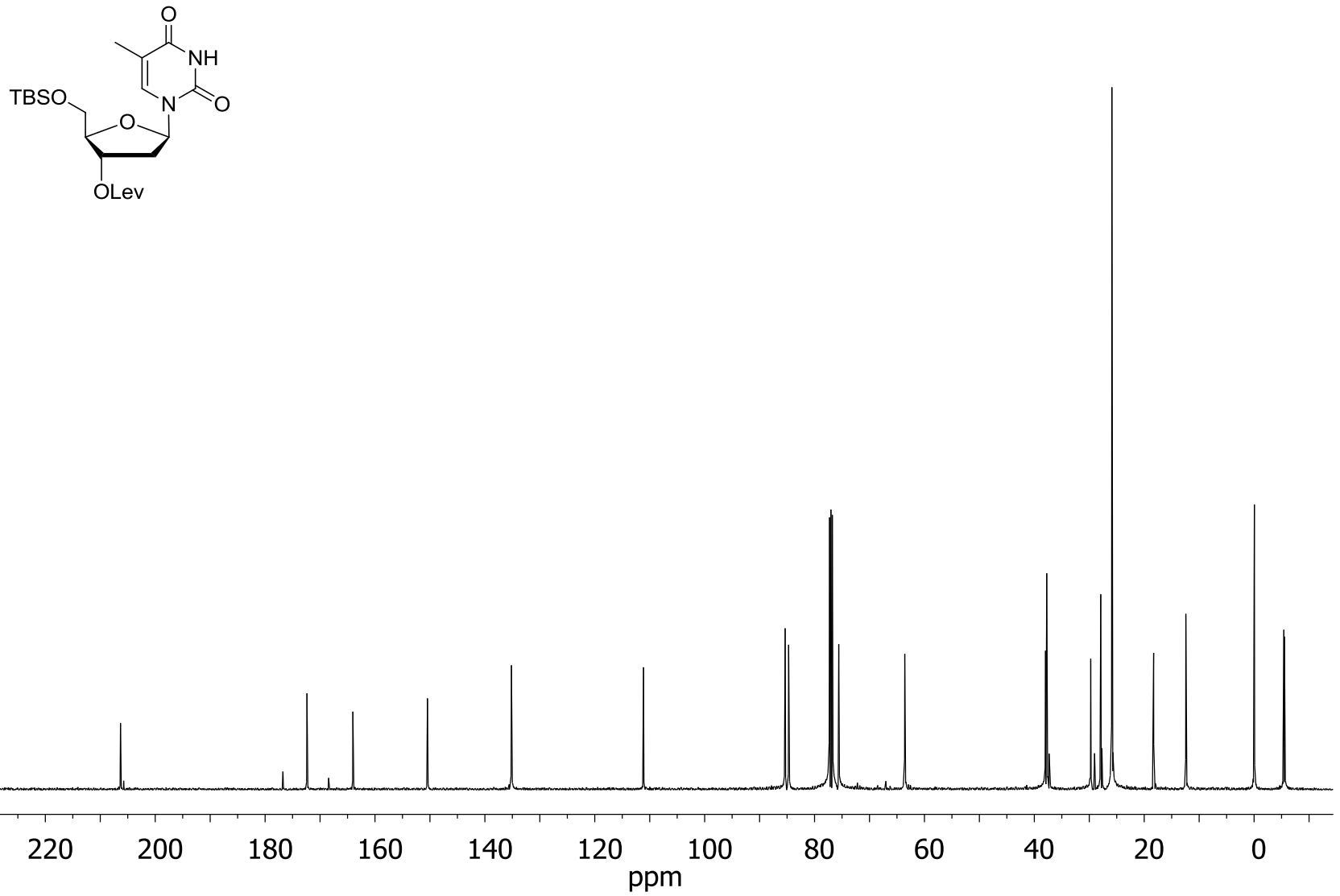


Supplementary Scheme 3: Synthesis of O^4-dG-alkylene-O^4-dT dimer phosphoramidites **16a** and **16b**. Reagents and conditions: (i) 1. TBS-Cl (1.05 eq), Py, 0°C →22°C, 16h, 2. Levulinic acid (2 eq), EDAC • HCl (2 eq), DMAP (cat) dioxane, 22°C, 16h. (ii) **4a** or **4b** (1.2 eq), Ph_3P (2.2 eq), DIAD (2.6 eq), dioxane, 22°C, 18h. (iii) N_2H_4 • H_2O (0.5 M) in pyridinium acetate buffer (1:1 v / v), 21°C, 30 min. (iv) Cl-P(OCEt)(N*i*Pr$_2$) (1.5 eq), DIPEA (1.9 eq), THF, 21°C, 30 min. It should be noted that the synthesis of compounds **4a** and **4b** has been reported previously.[2]

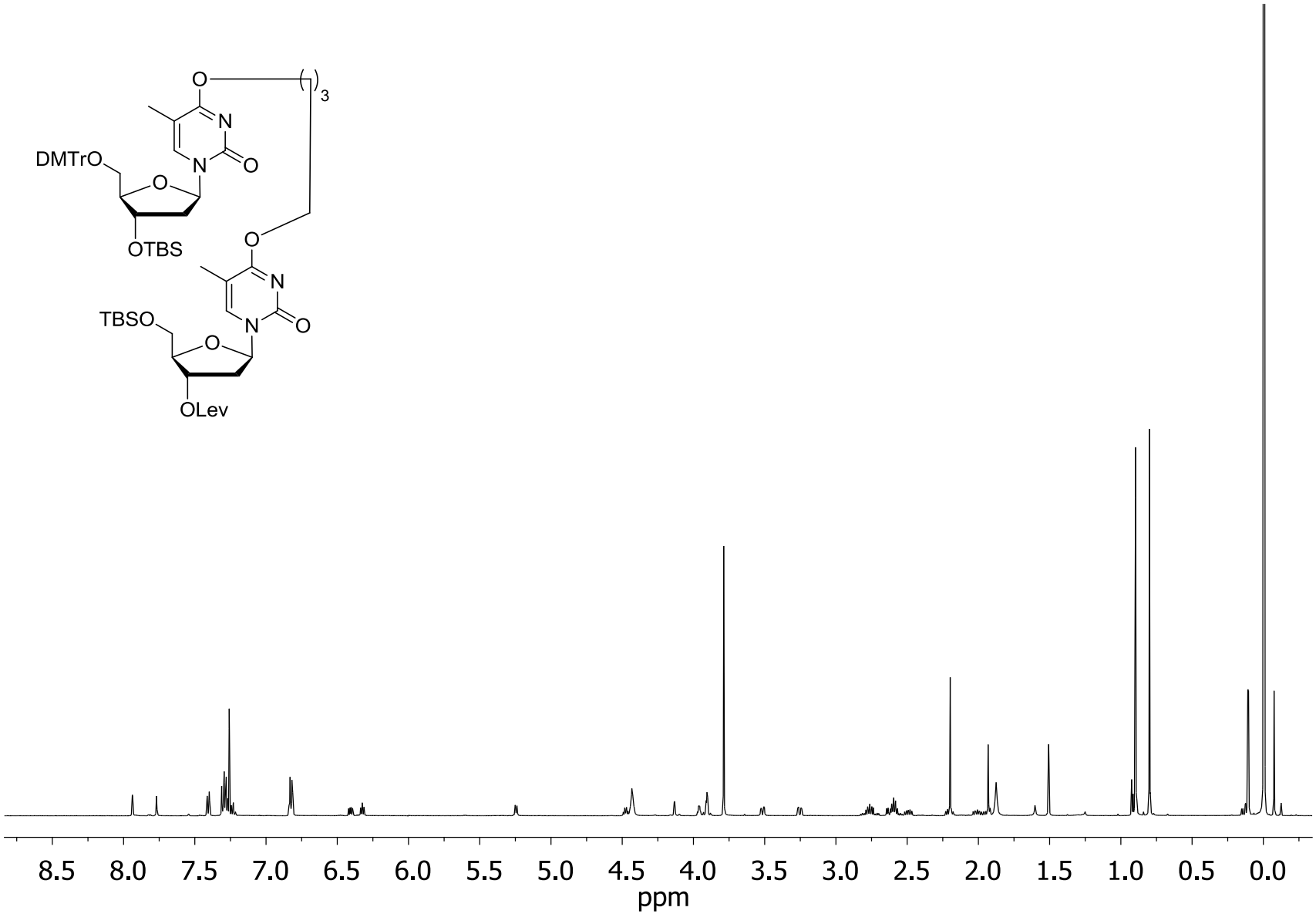
Supplementary Figure 1 - 500 MHz ^{1}H NMR spectrum of compound (**2**) (in $CDCl_3$)



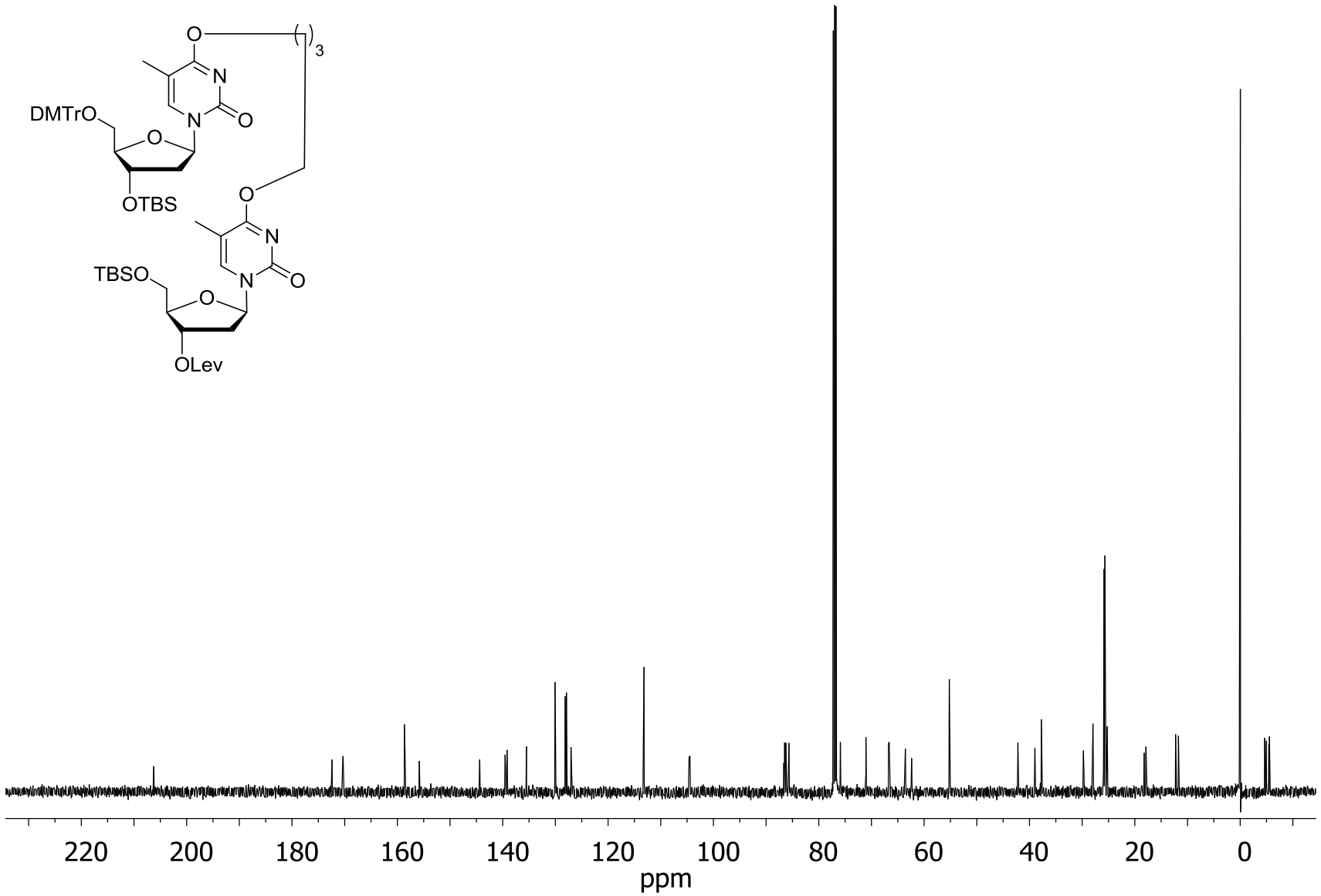
Supplementary Figure 2 - 125.7 MHz ^{13}C NMR spectrum of compound (**2**) (in $CDCl_3$)



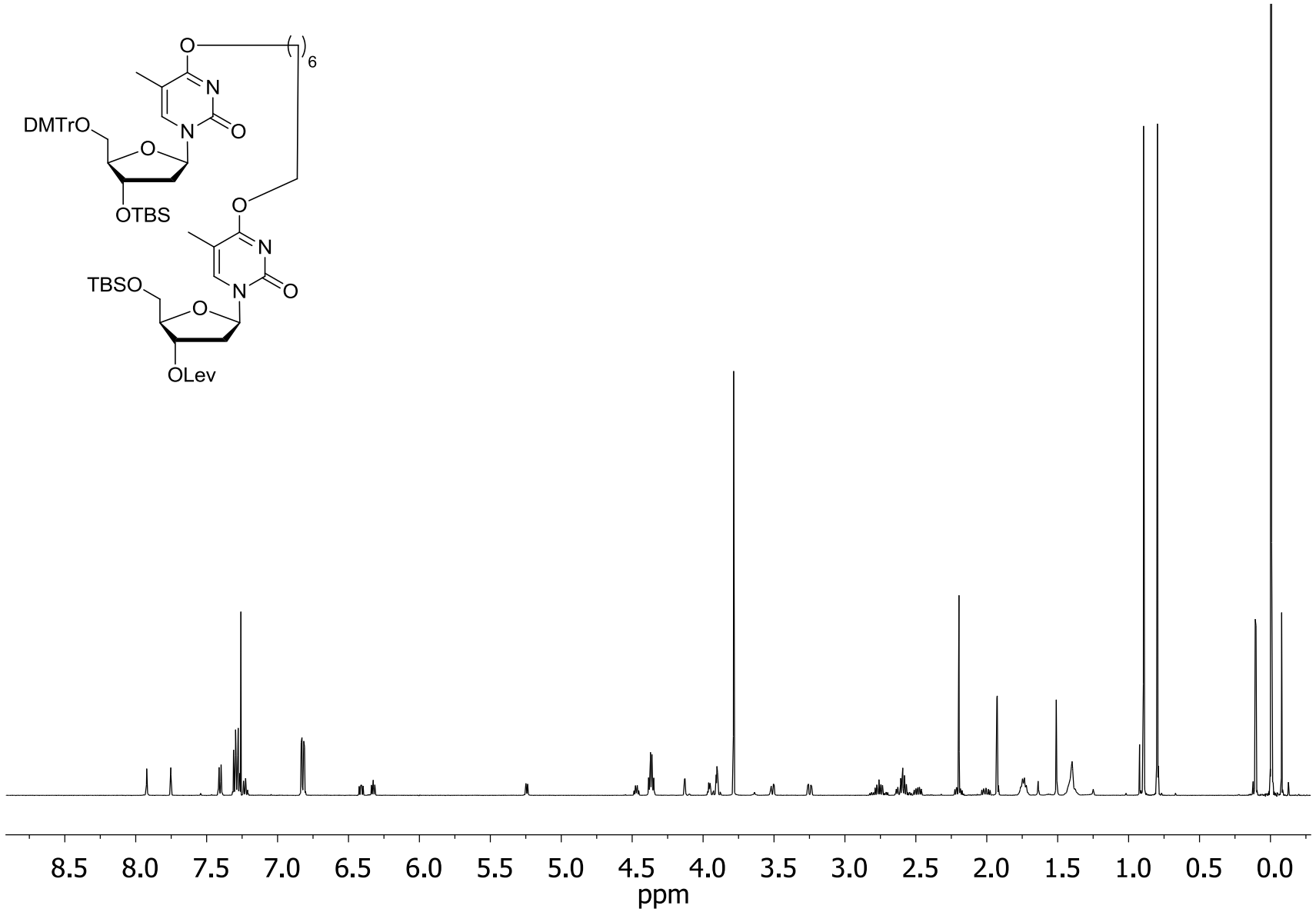
Supplementary Figure 3 - 500 MHz ^{1}H NMR spectrum of compound (**5a**) (in $CDCl_3$)



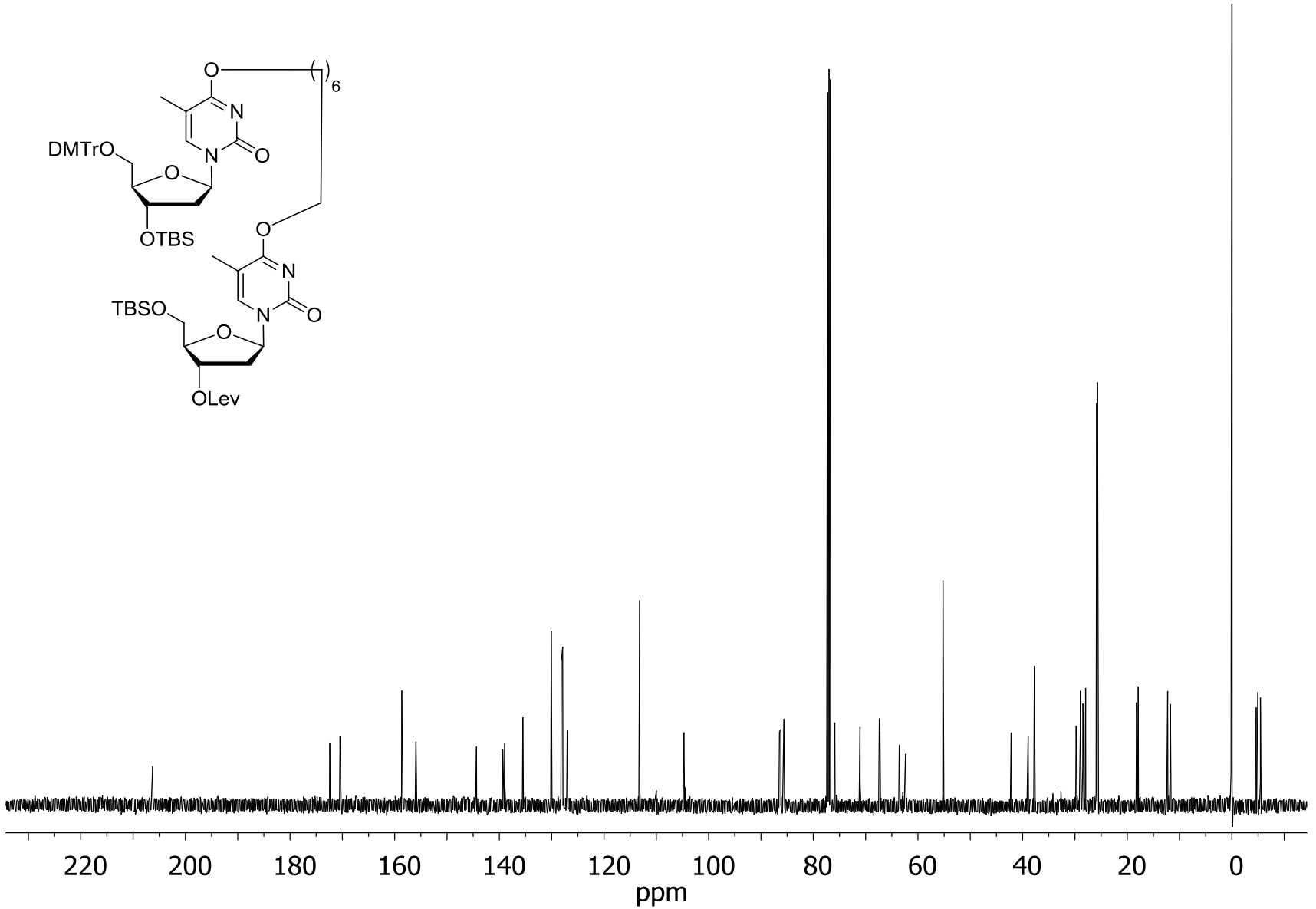
Supplementary Figure 4- 125.7 MHz ^{13}C NMR spectrum of compound (**5a**) (in $CDCl_3$)



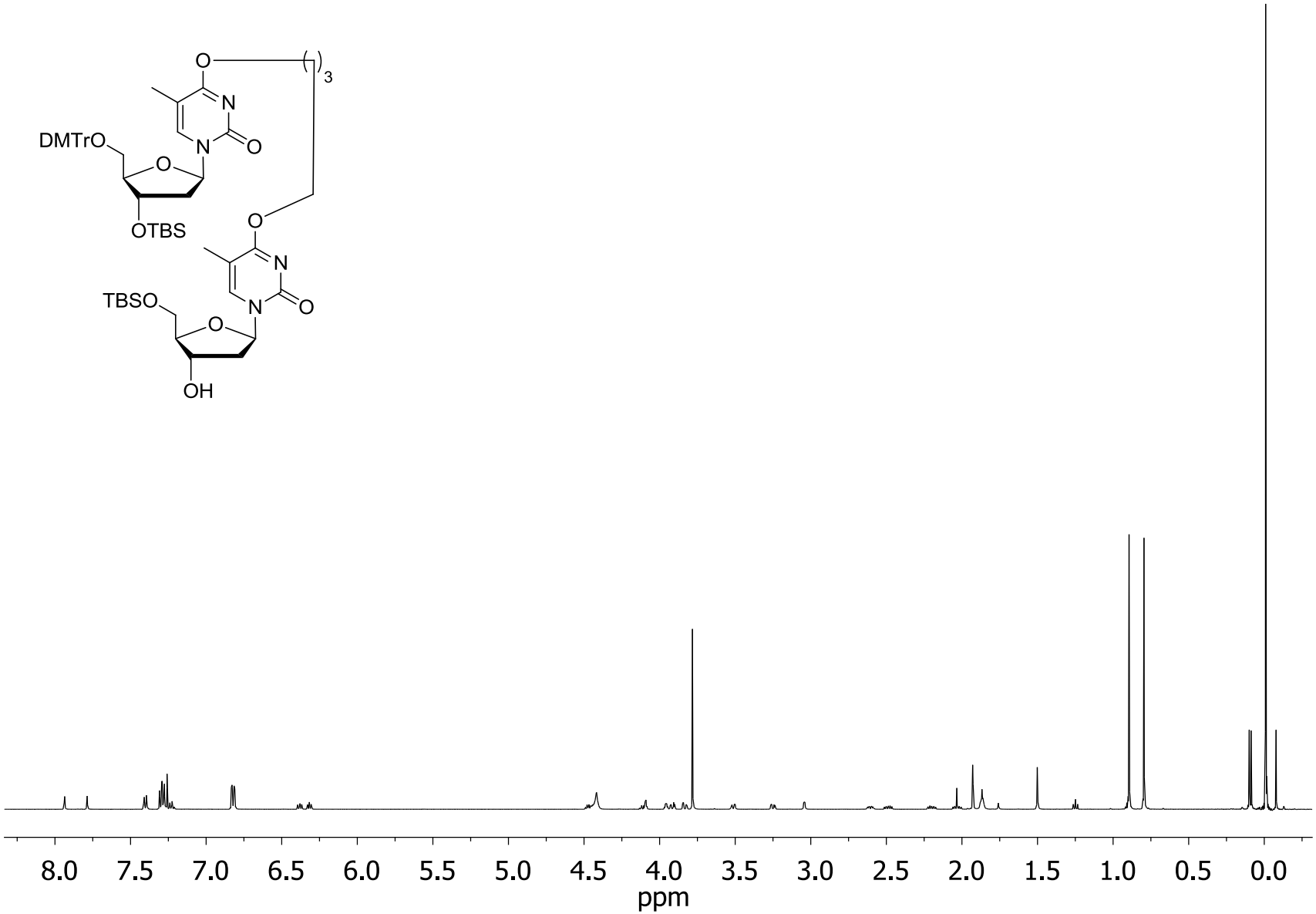
Supplementary Figure 5 - 500 MHz ^{1}H NMR spectrum of compound (**5b**) (in $CDCl_3$)



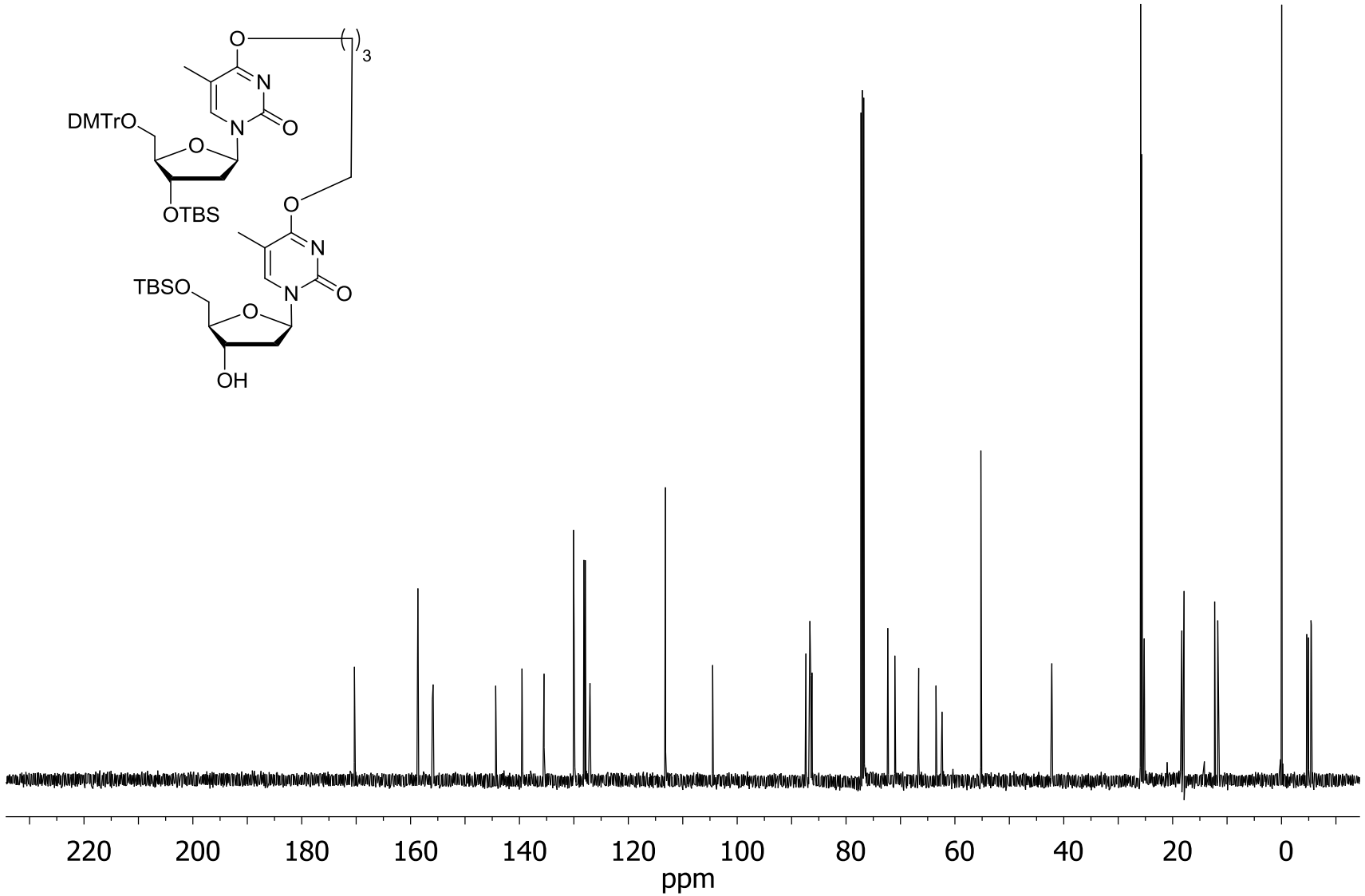
Supplementary Figure 6 - 125.7 MHz ^{13}C NMR spectrum of compound (**5b**) (in $CDCl_3$)



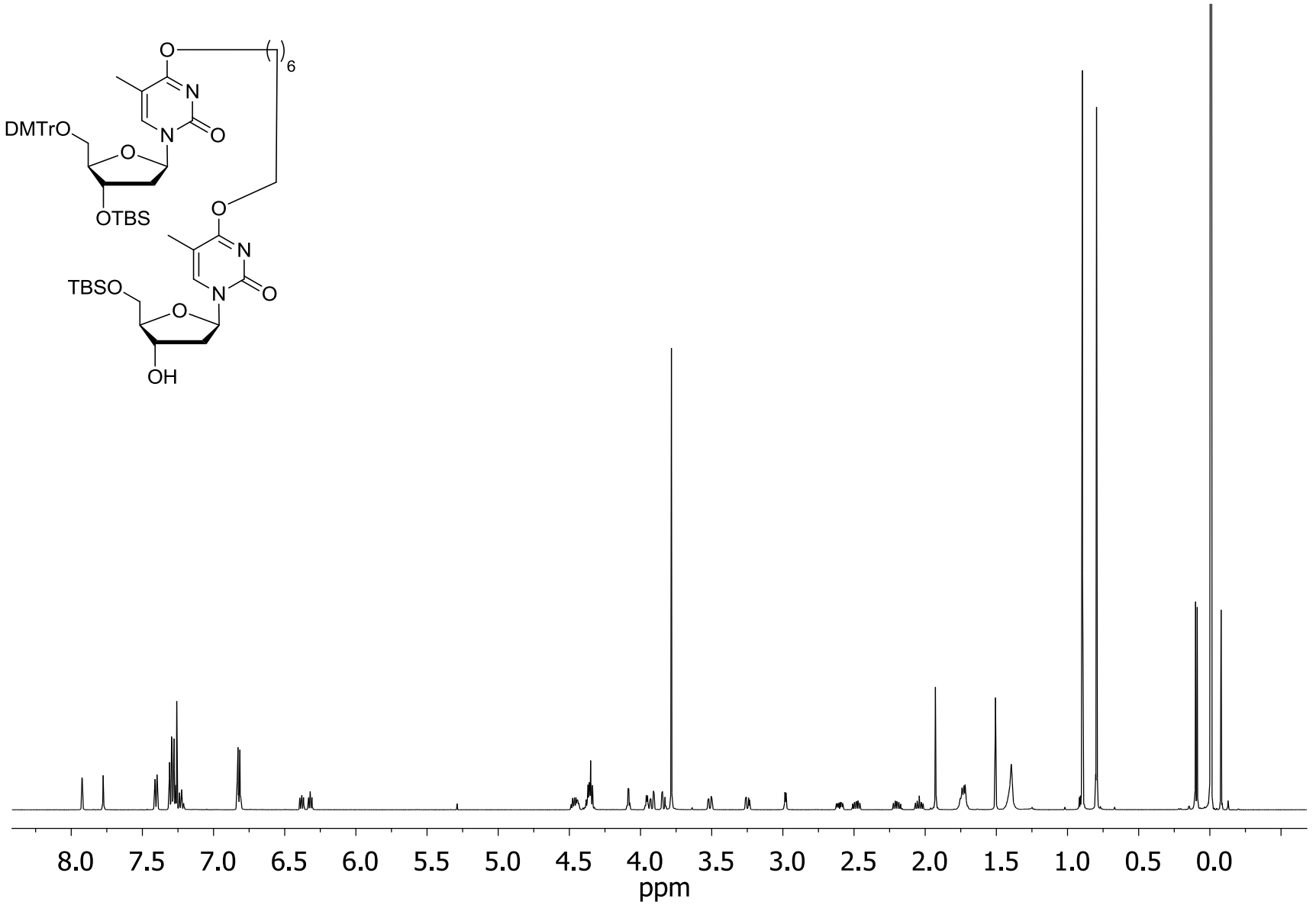
Supplementary Figure 7 - 500 MHz ^{1}H NMR spectrum of compound (**6a**) (in $CDCl_3$)



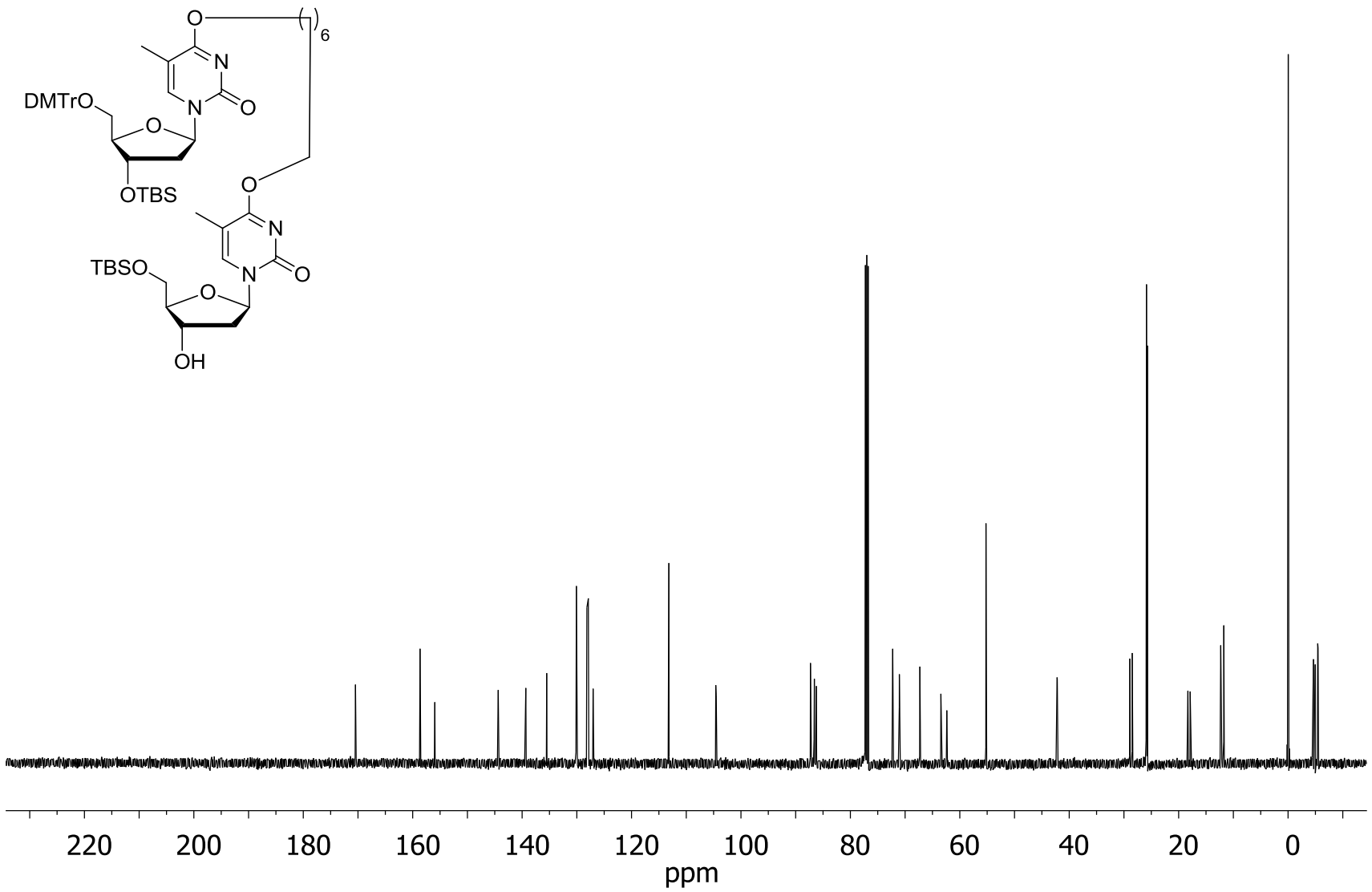
Supplementary Figure 8 - 125.7 MHz ^{13}C NMR spectrum of compound (**6a**) (in $CDCl_3$)



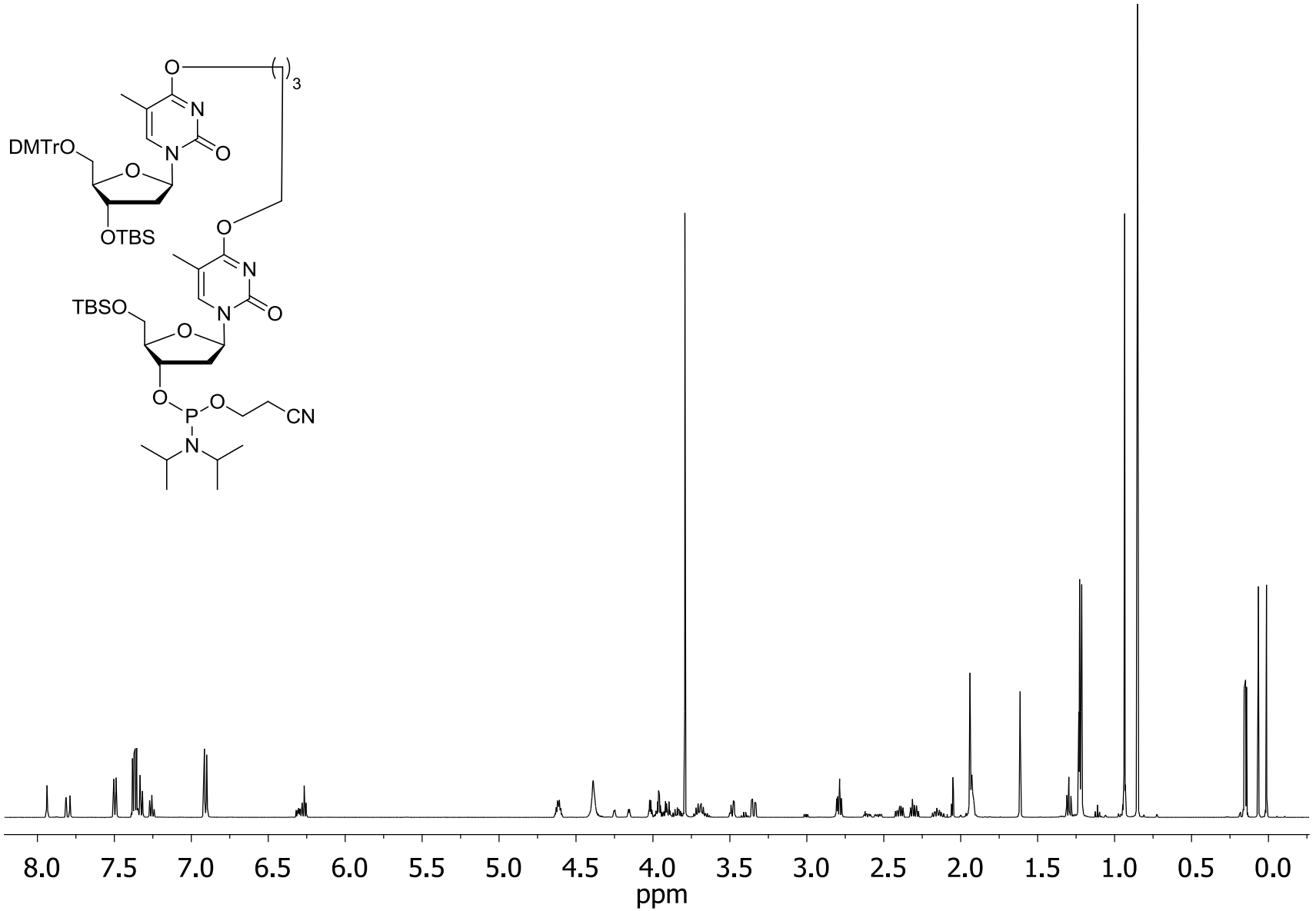
Supplementary Figure 9 - 500 MHz ^{1}H NMR spectrum of compound (**6b**) (in $CDCl_3$)



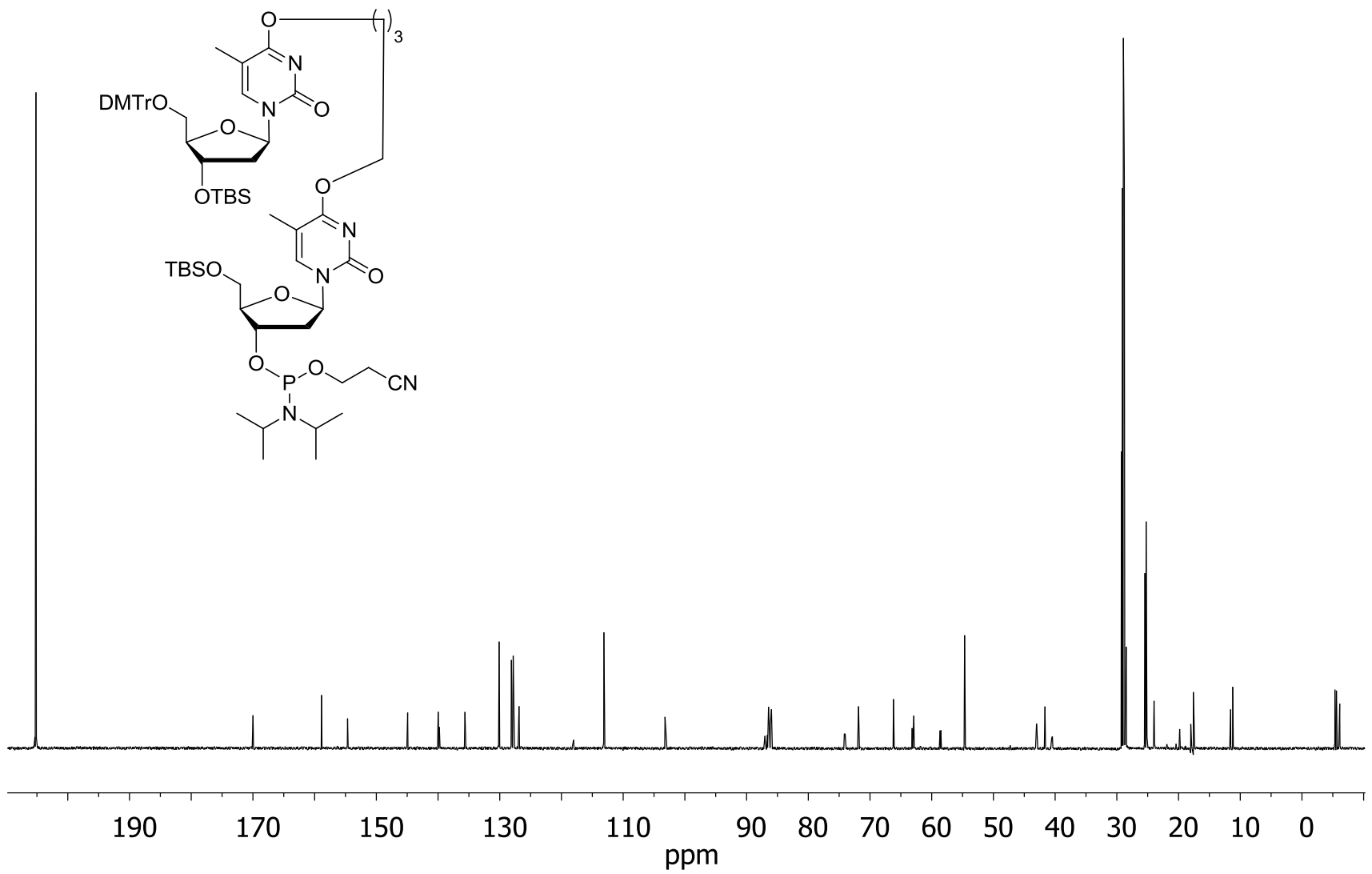
Supplementary Figure 10 - 125.7 MHz ^{13}C NMR spectrum of compound (**6b**) (in $CDCl_3$)



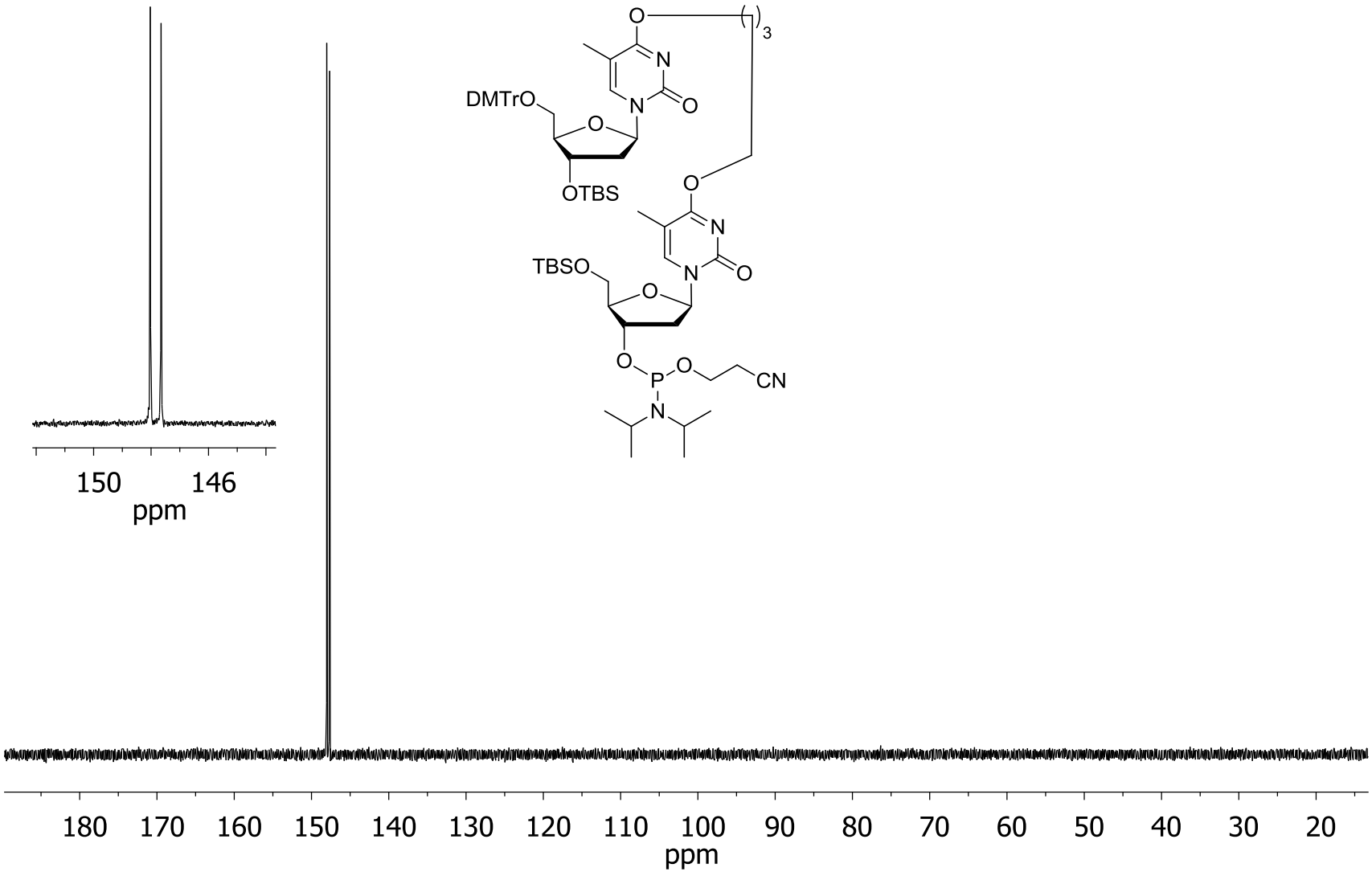
Supplementary Figure 11 - 500 MHz ^{1}H NMR spectrum of compound (**7a**) (in d_6-acetone)



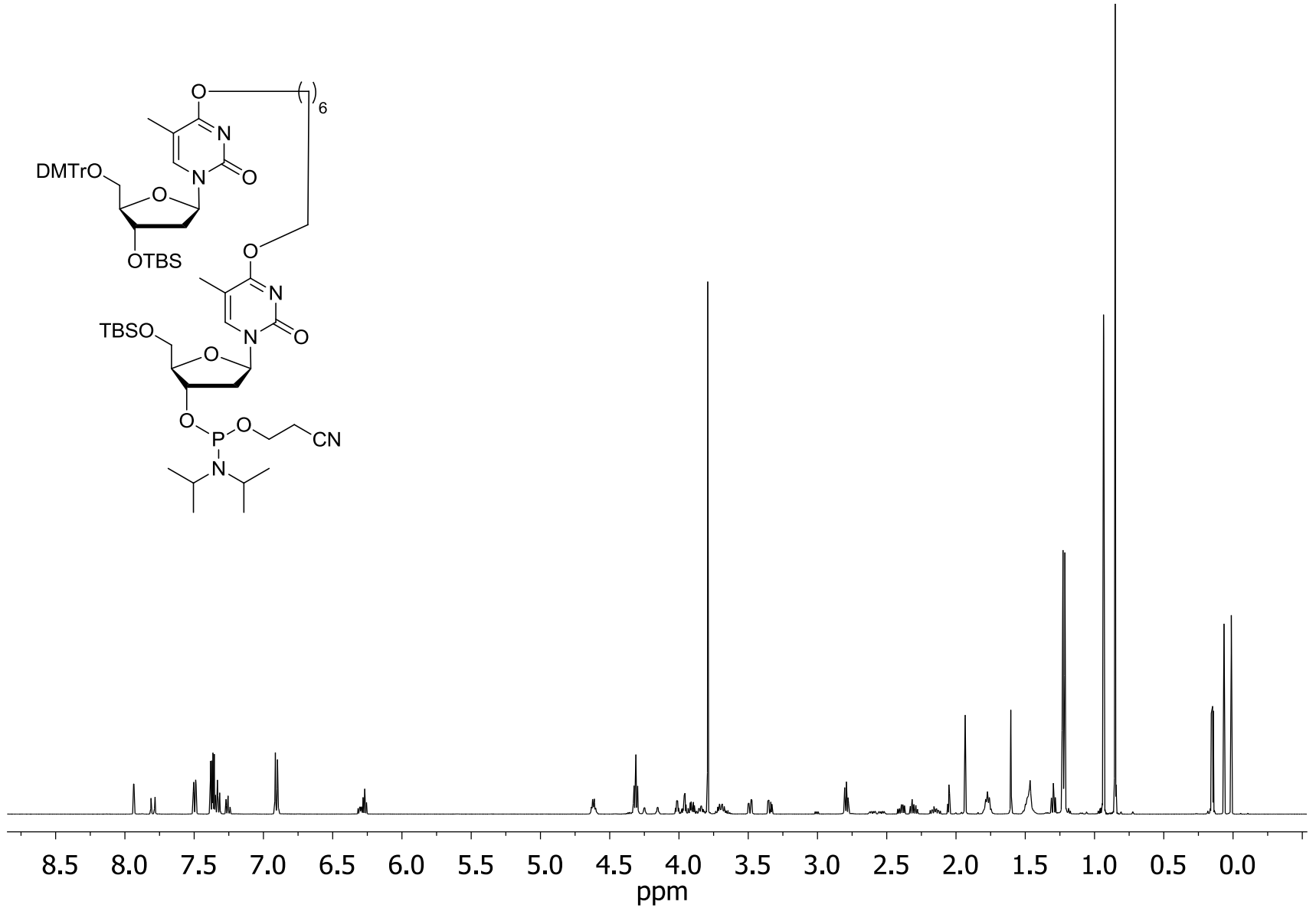
Supplementary Figure 12 - 125.7 MHz ^{13}C NMR spectrum of compound (**7a**) (in d_6-acetone)



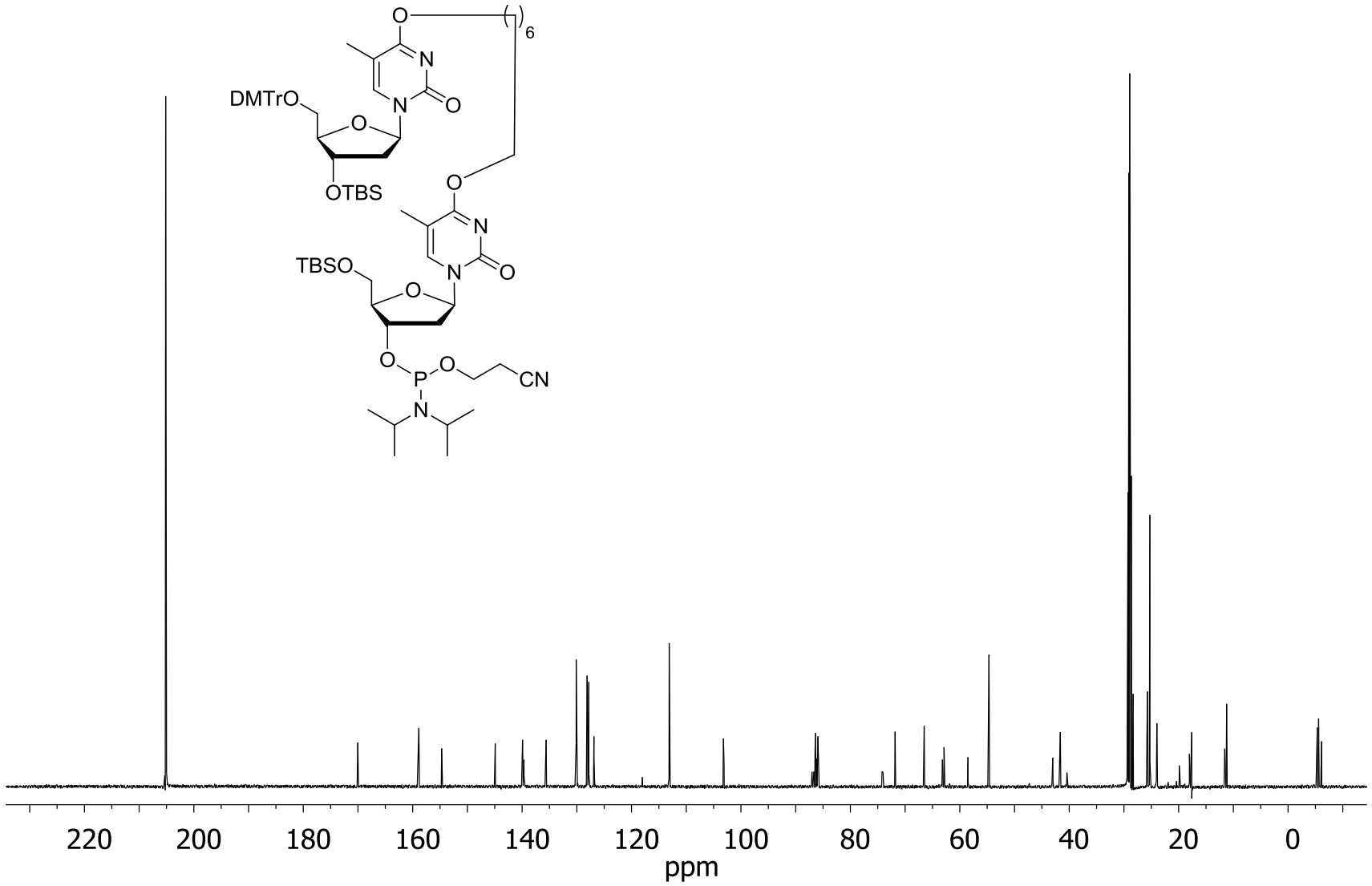
Supplementary Figure 13 - 202.3 MHz ^{31}P NMR spectrum of compound (**7a**) (d_6-acetone)



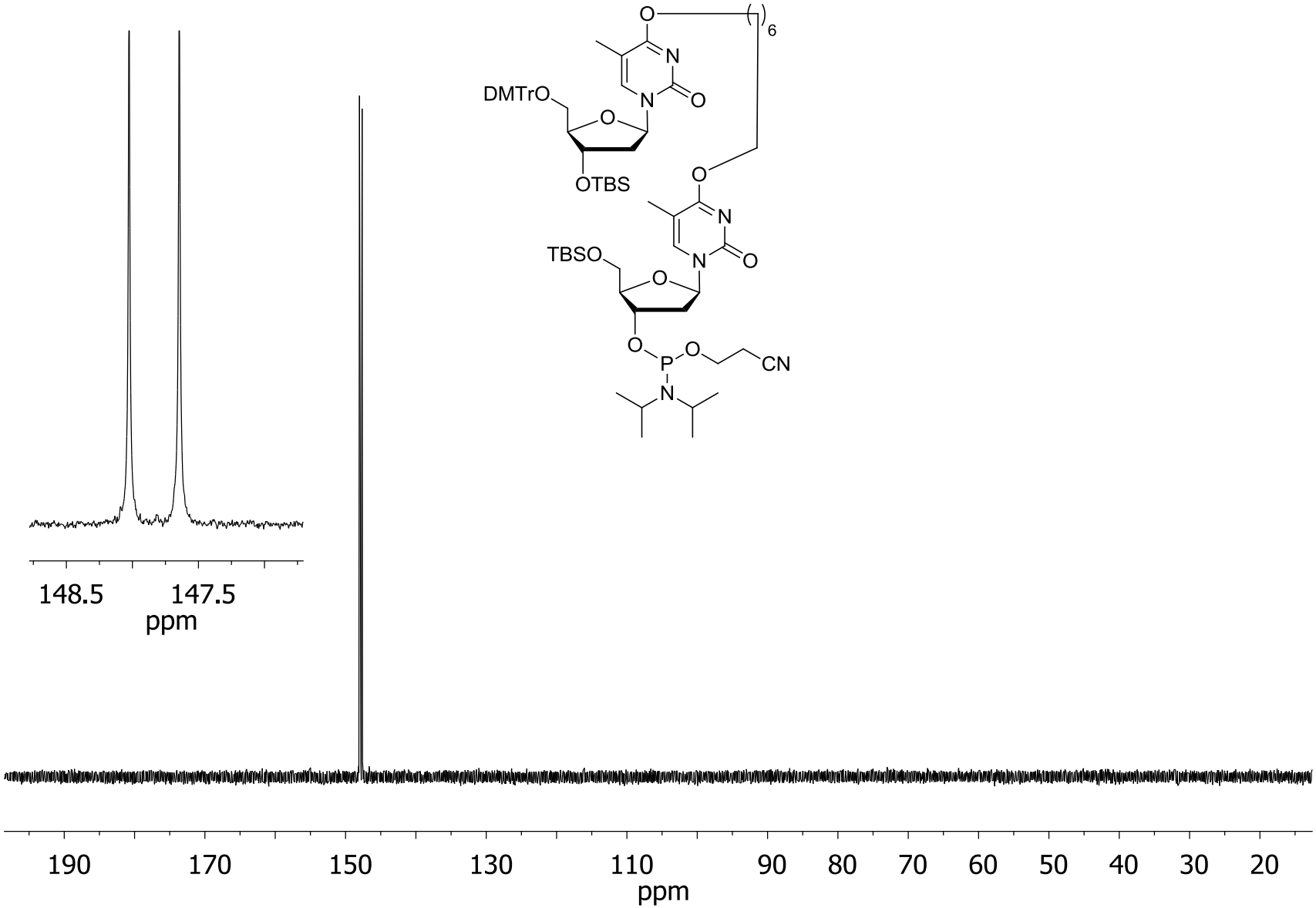
Supplementary Figure 14 - 500 MHz ^{1}H NMR spectrum of compound (**7b**) (in d_6-acetone)



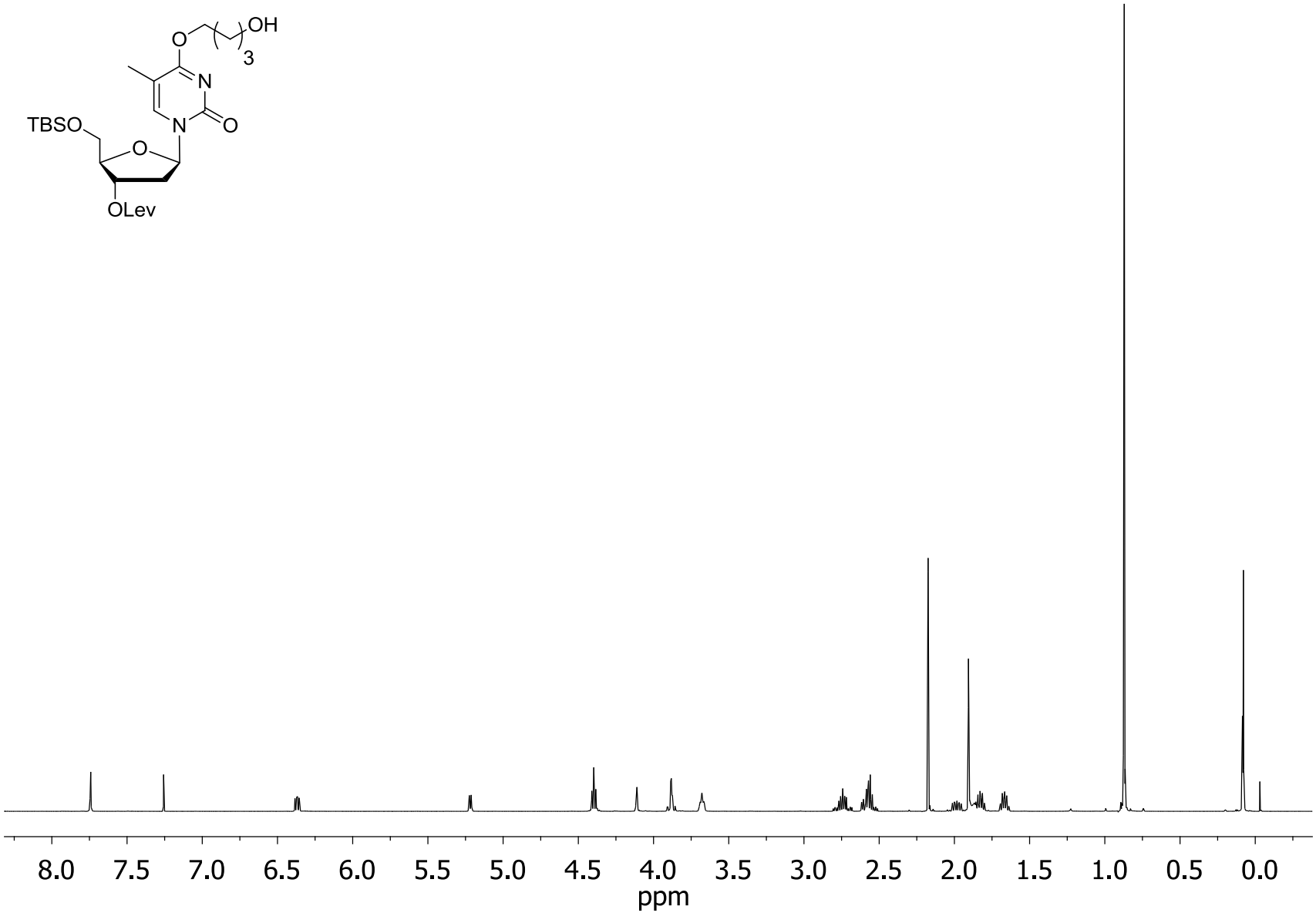
Supplementary Figure 15 - 125.7 MHz ^{13}C NMR spectrum of compound (**7b**) (in d_6-acetone)



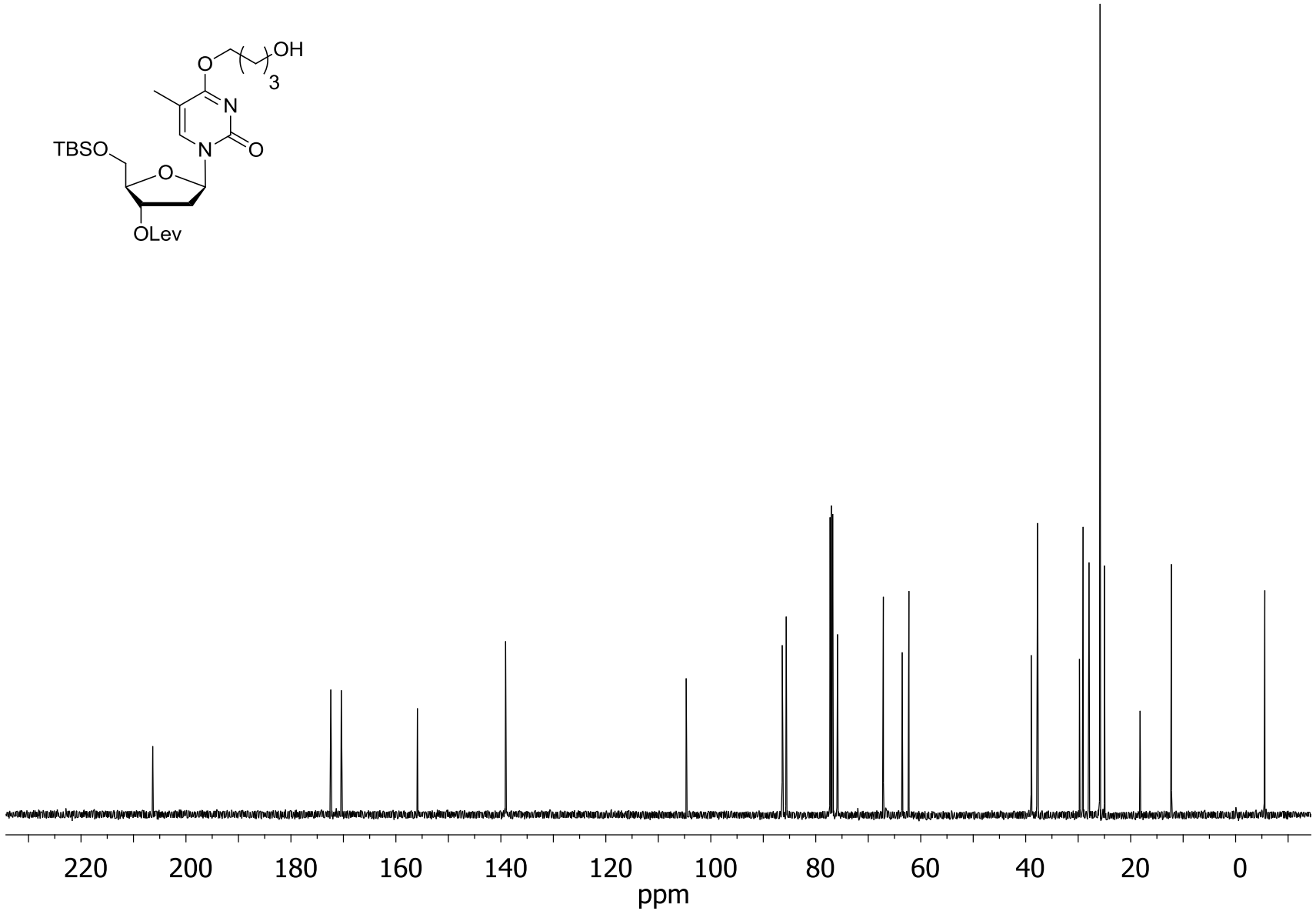
Supplementary Figure 16 - 202.3 MHz ^{31}P NMR spectrum of compound (**7b**) (in d_6-acetone)



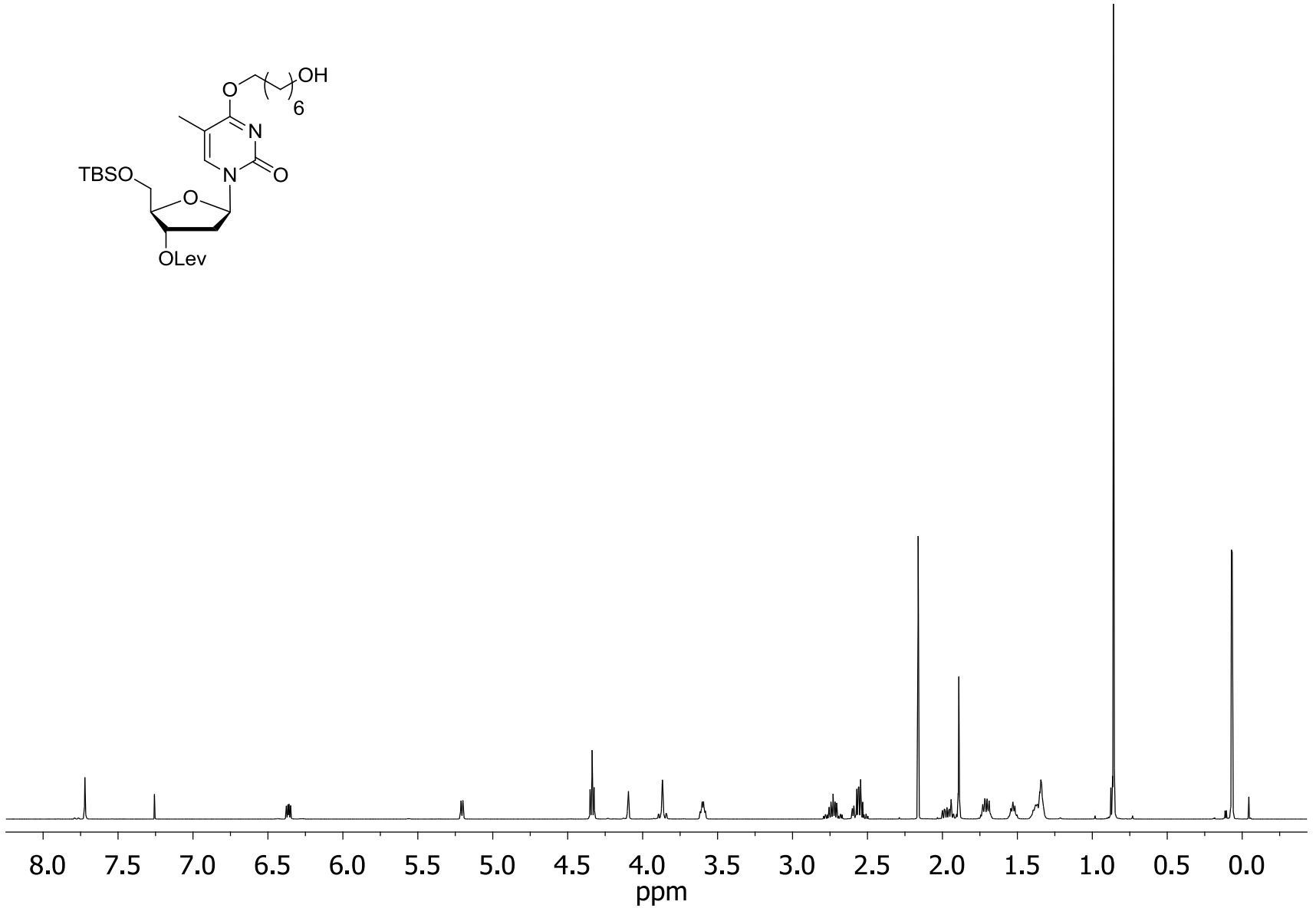
Supplementary Figure 17 - 500 MHz ^{1}H NMR spectrum of compound (**8a**) (in $CDCl_3$)



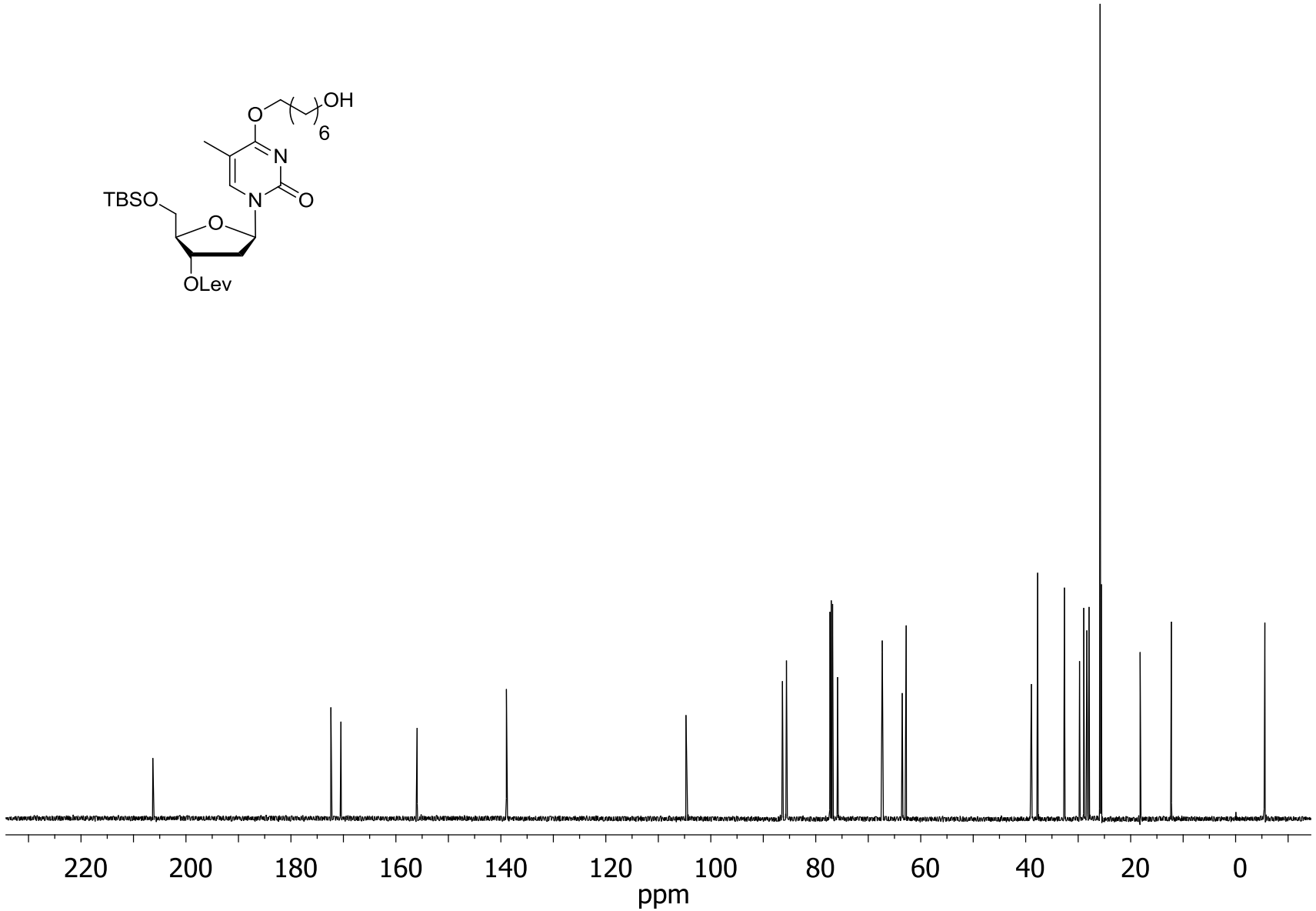
Supplementary Figure 18 - 125.7 MHz ^{13}C NMR spectrum of compound (**8a**) (in $CDCl_3$)



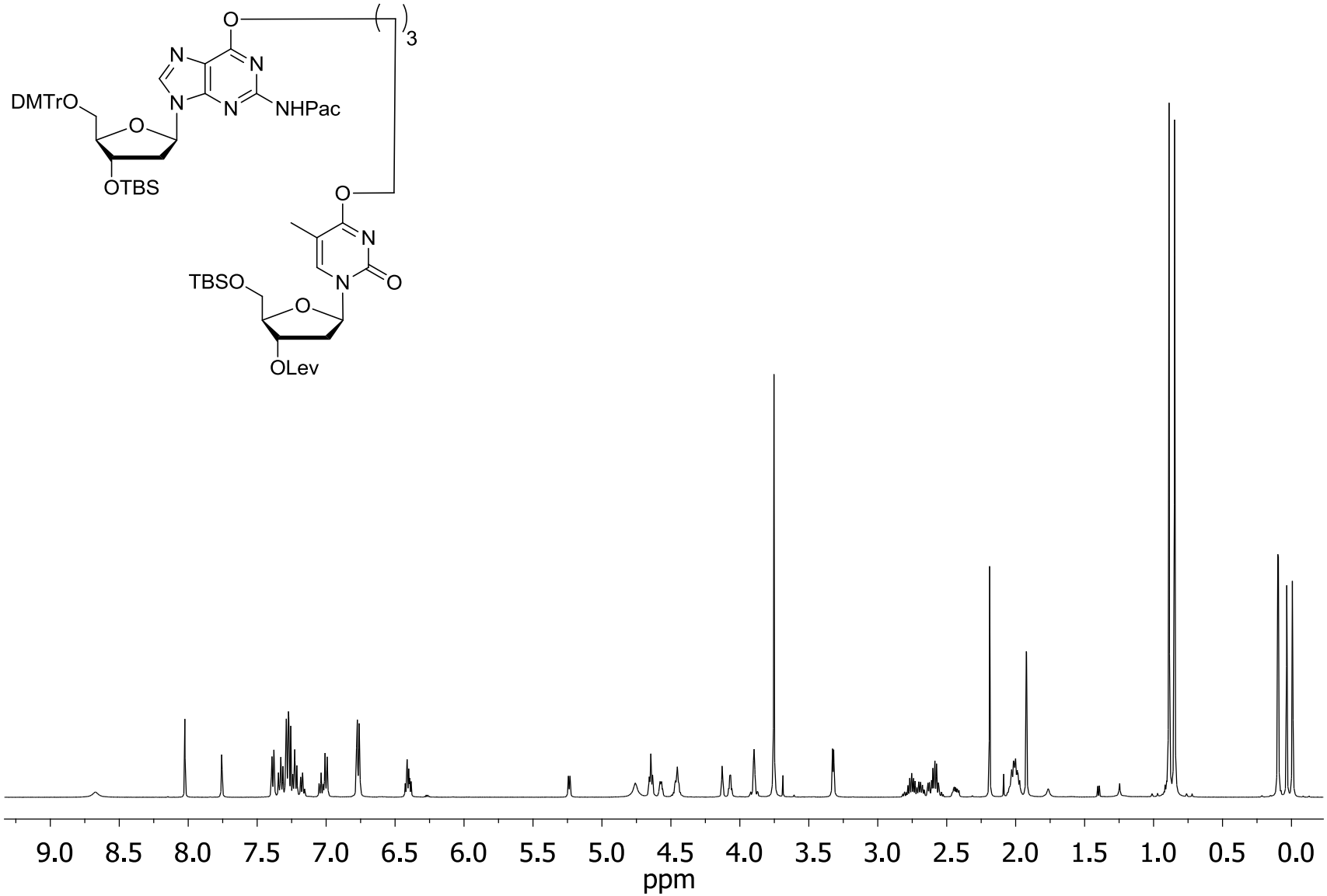
Supplementary Figure 19 - 500 MHz ^{1}H NMR spectrum of compound (**8b**) (in $CDCl_3$)



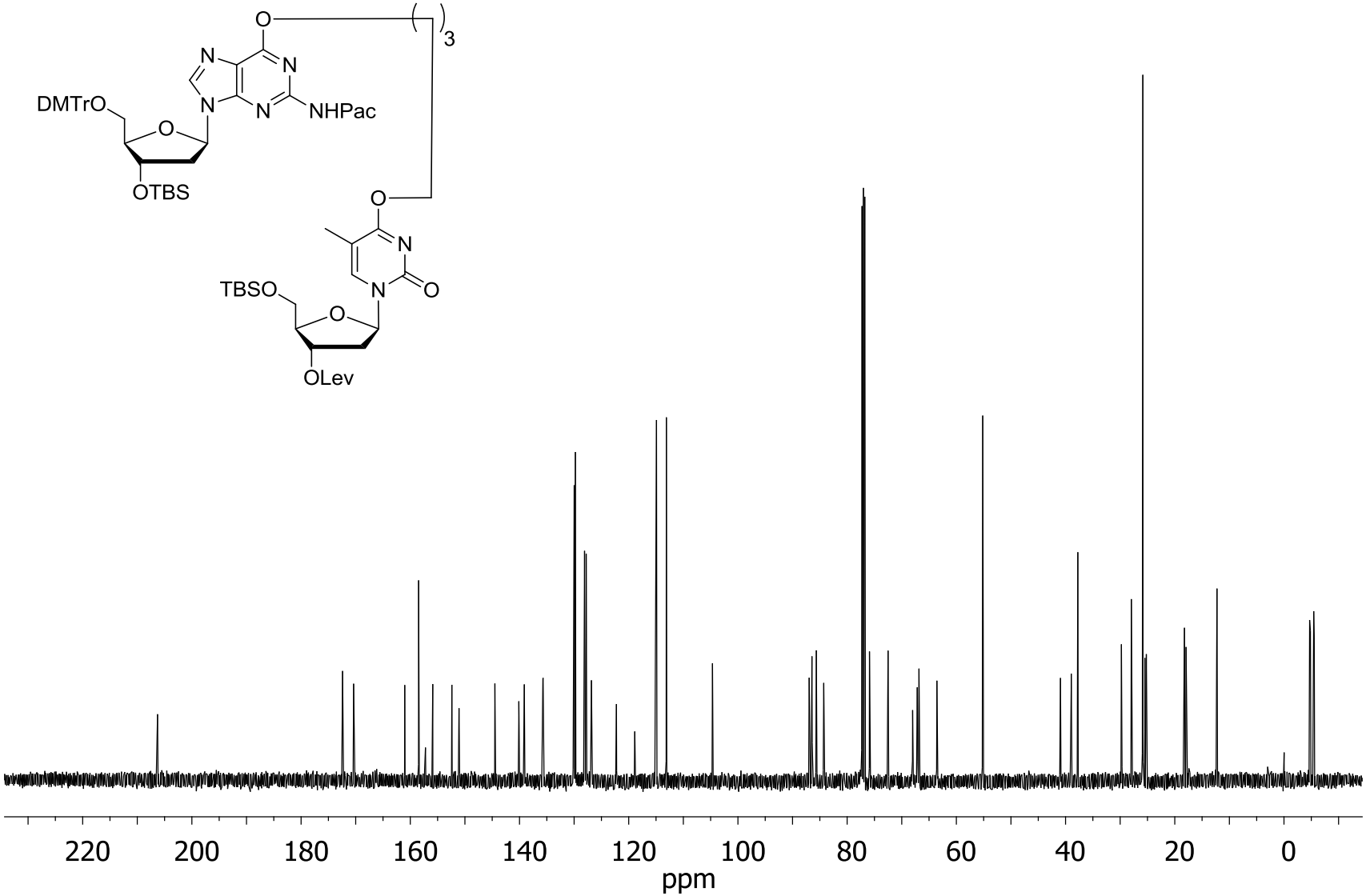
Supplementary Figure 20 - 125.7 MHz ^{13}C NMR spectrum of compound (**8b**) (in $CDCl_3$)



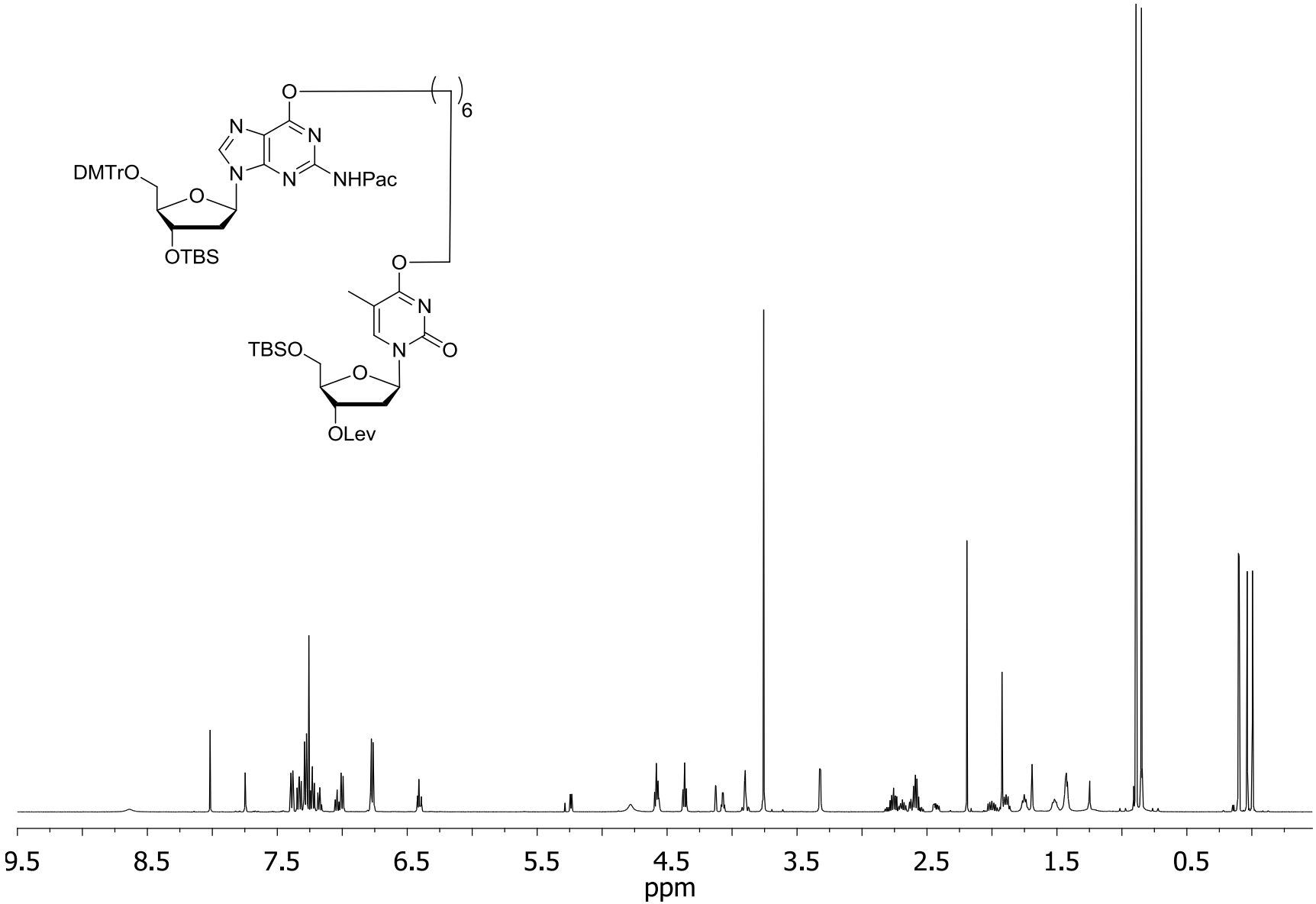
Supplementary Figure 21 - 500 MHz ^{1}H NMR spectrum of compound (**9a**) (in $CDCl_3$)



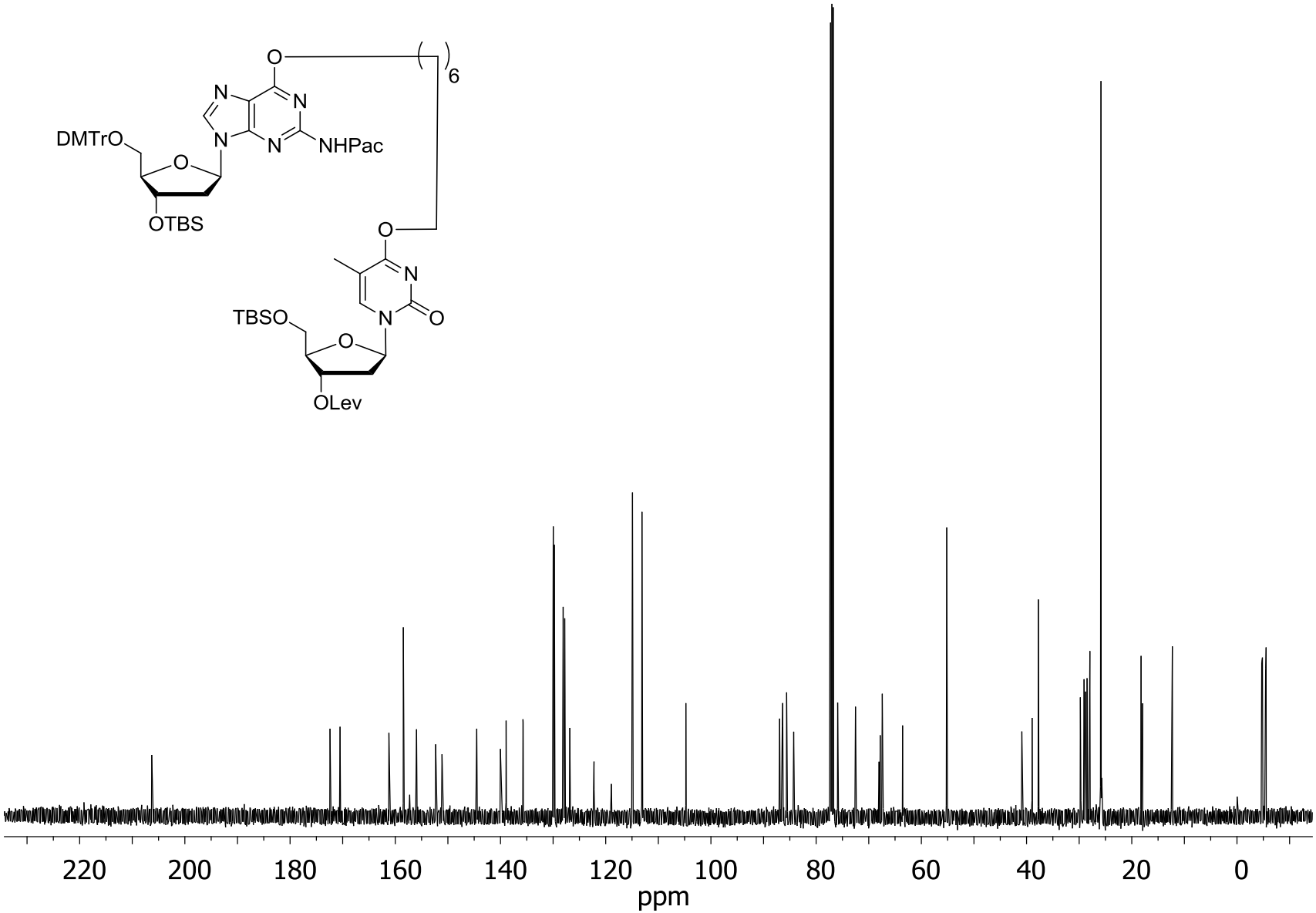
Supplementary Figure 22 - 125.7 MHz ^{13}C NMR spectrum of compound (**9a**) (in $CDCl_3$)



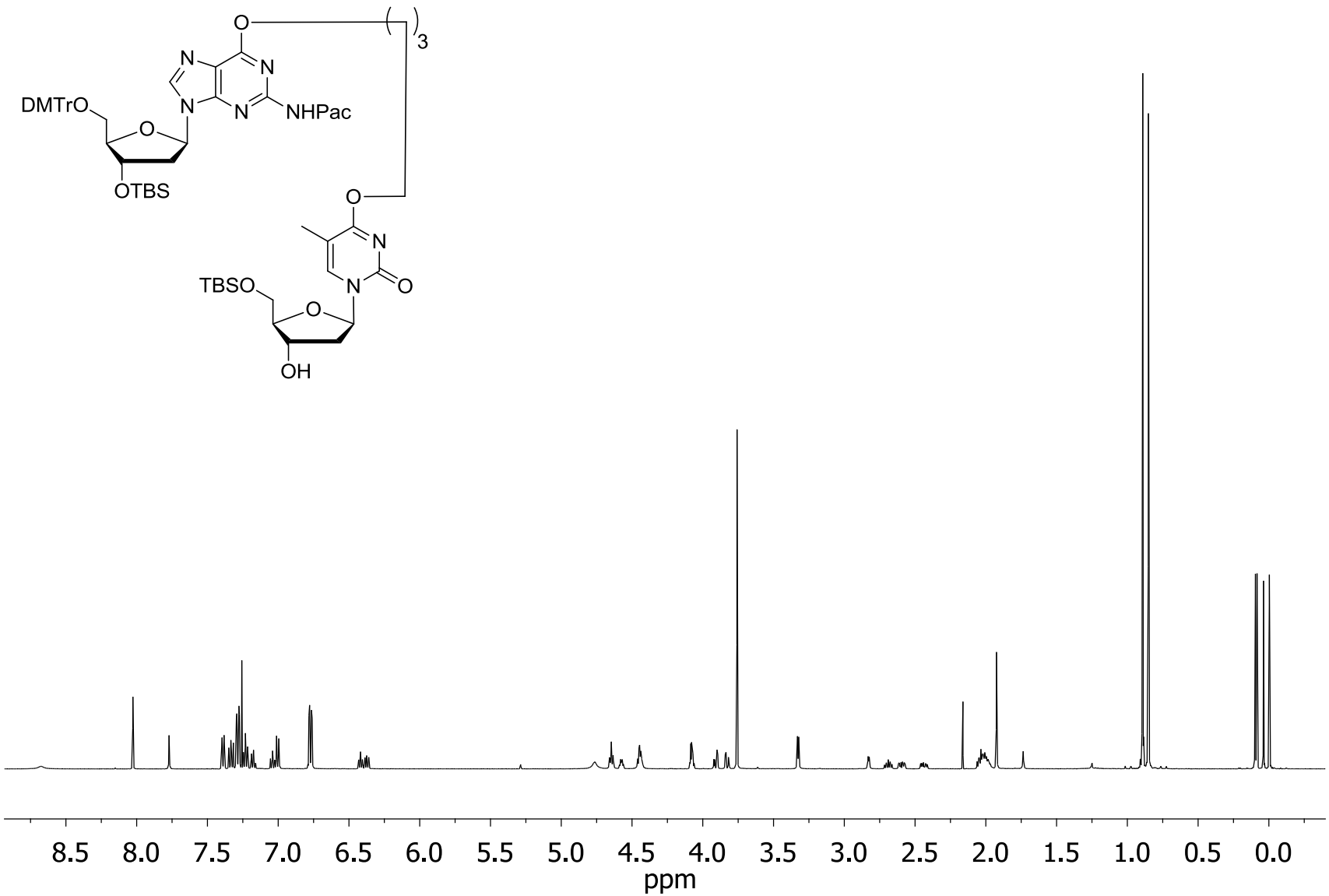
Supplementary Figure 23 - 500 MHz ^{1}H NMR spectrum of compound (**9b**) (in $CDCl_3$)



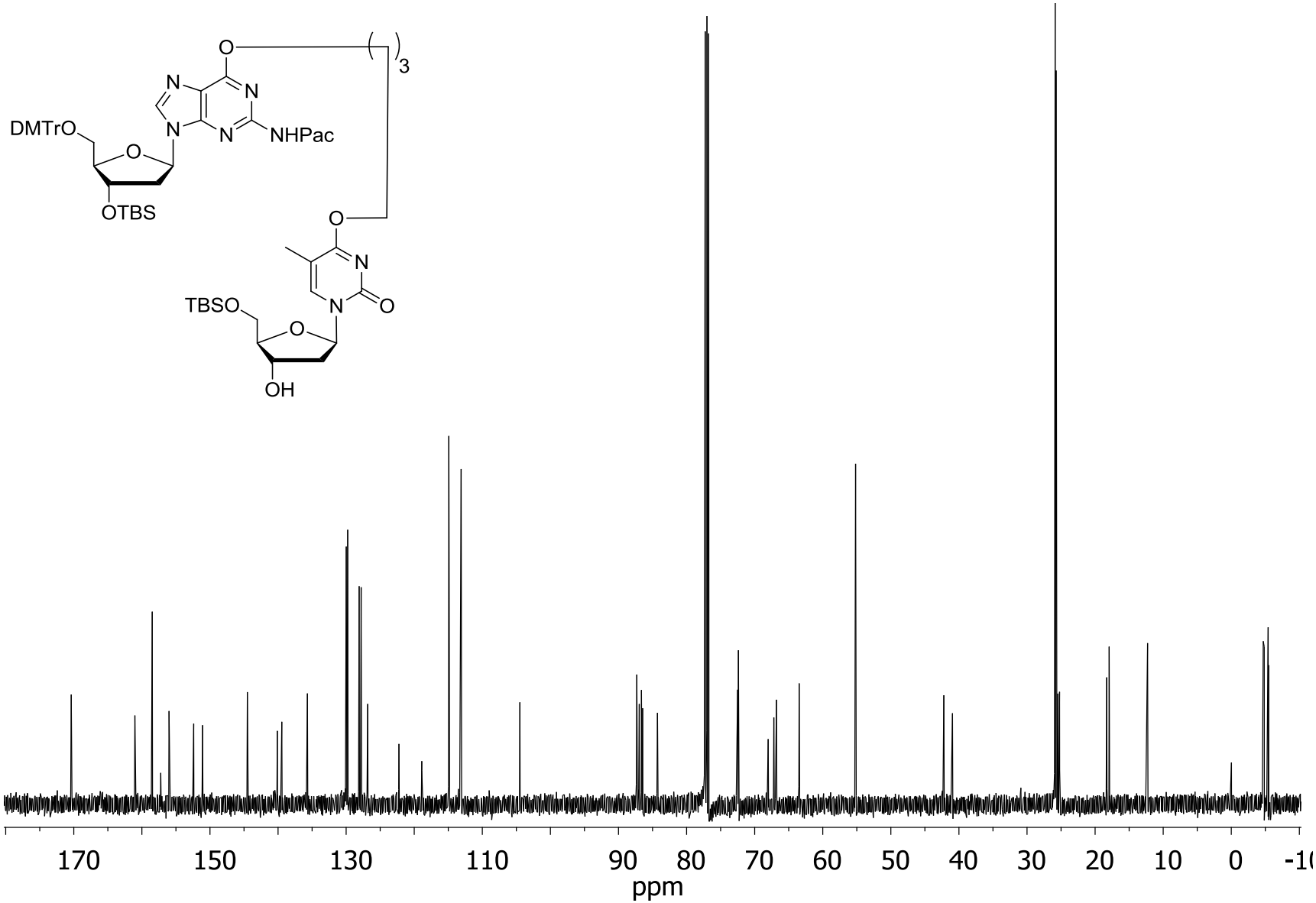
Supplementary Figure 24 - 125.7 MHz ^{13}C NMR spectrum of compound (**9b**) (in $CDCl_3$)



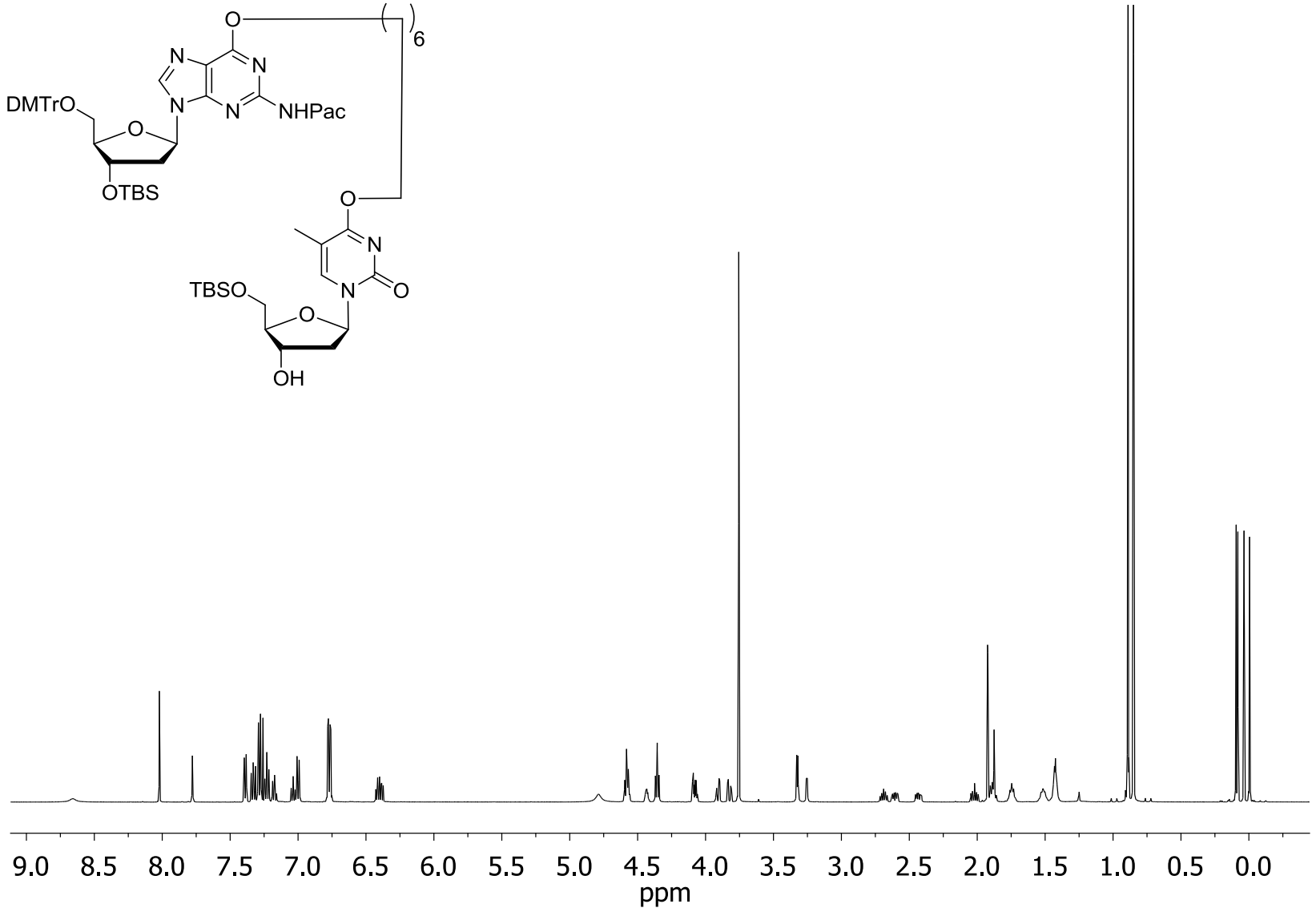
Supplementary Figure 25 - 500 MHz ^{1}H NMR spectrum of compound (**10a**) (in $CDCl_3$)



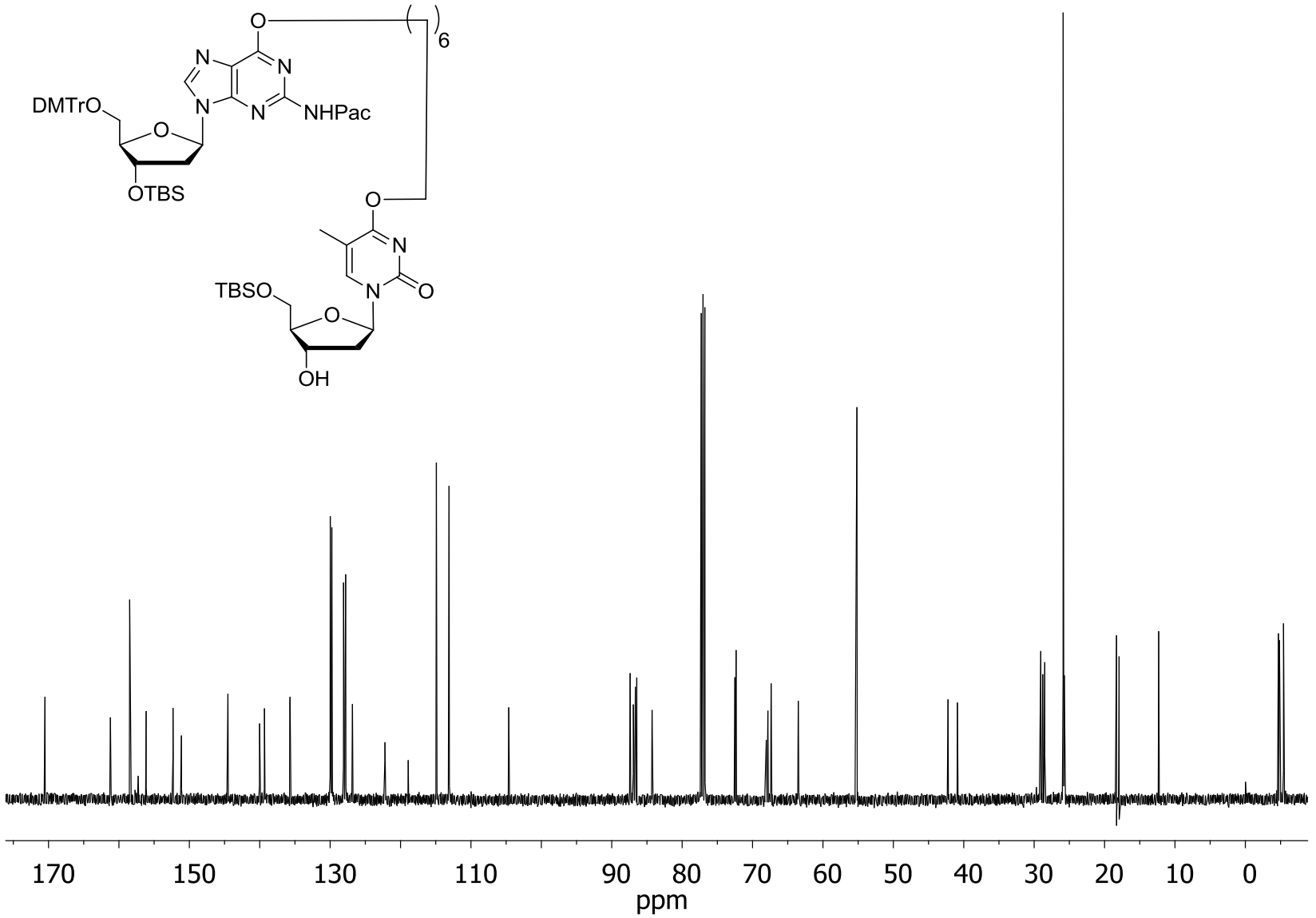
Supplementary Figure 26 - 125.7 MHz ^{13}C NMR spectrum of compound (**10a**) (in $CDCl_3$)



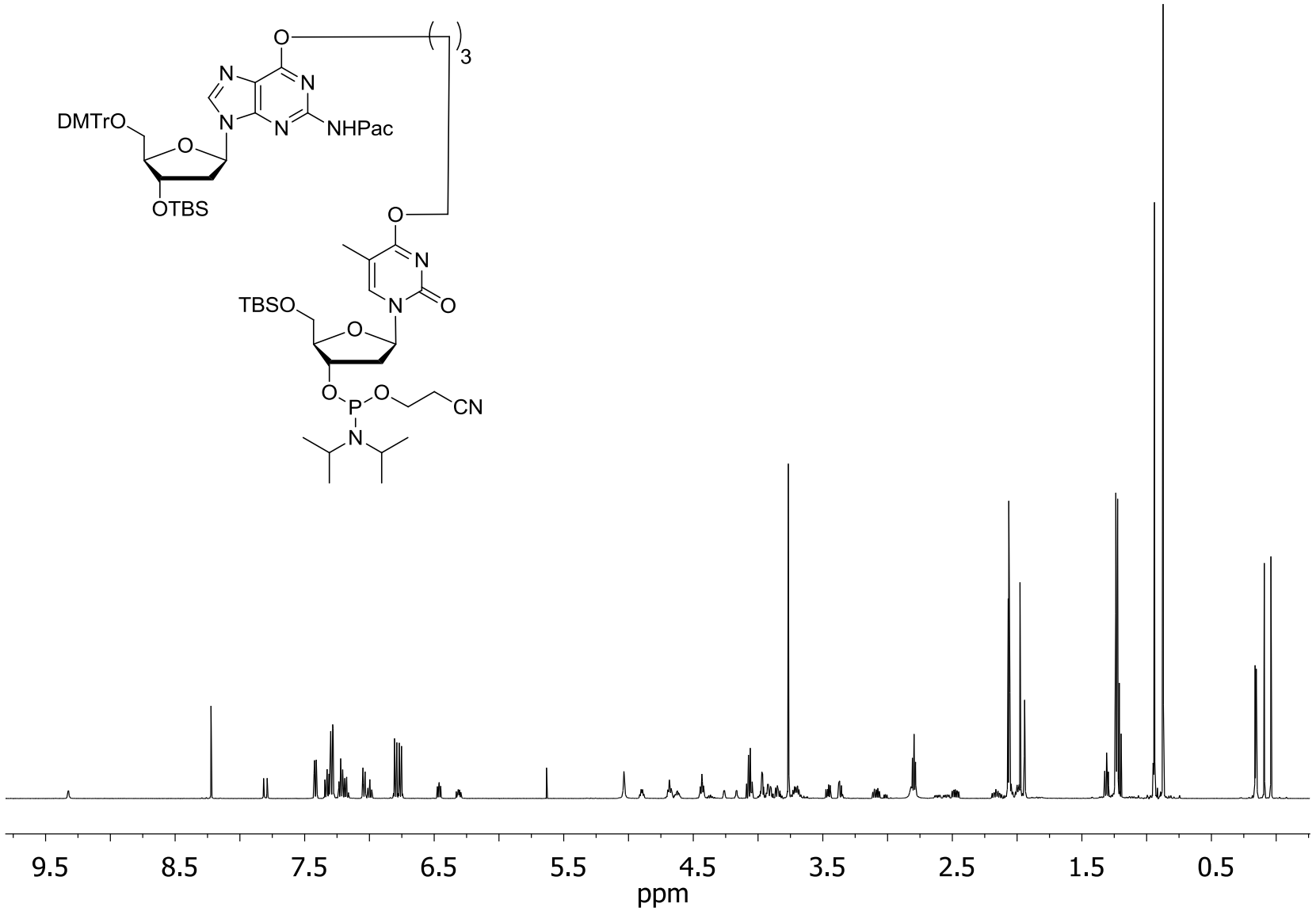
Supplementary Figure 27 - 500 MHz ^{1}H NMR spectrum of compound (**10b**) (in $CDCl_3$)



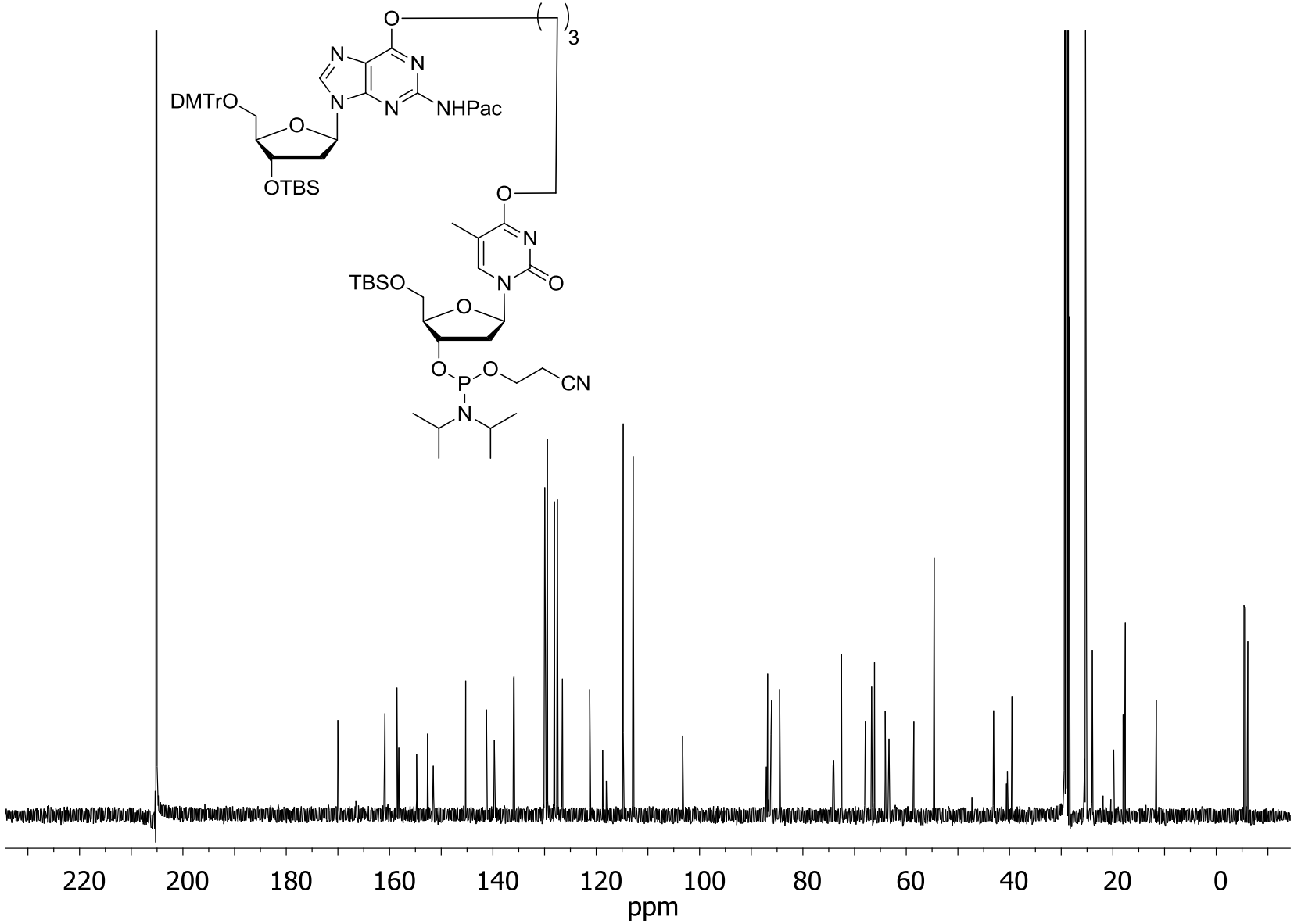
Supplementary Figure 28 - 125.7 MHz ^{13}C NMR spectrum of compound (**10b**) (in $CDCl_3$)



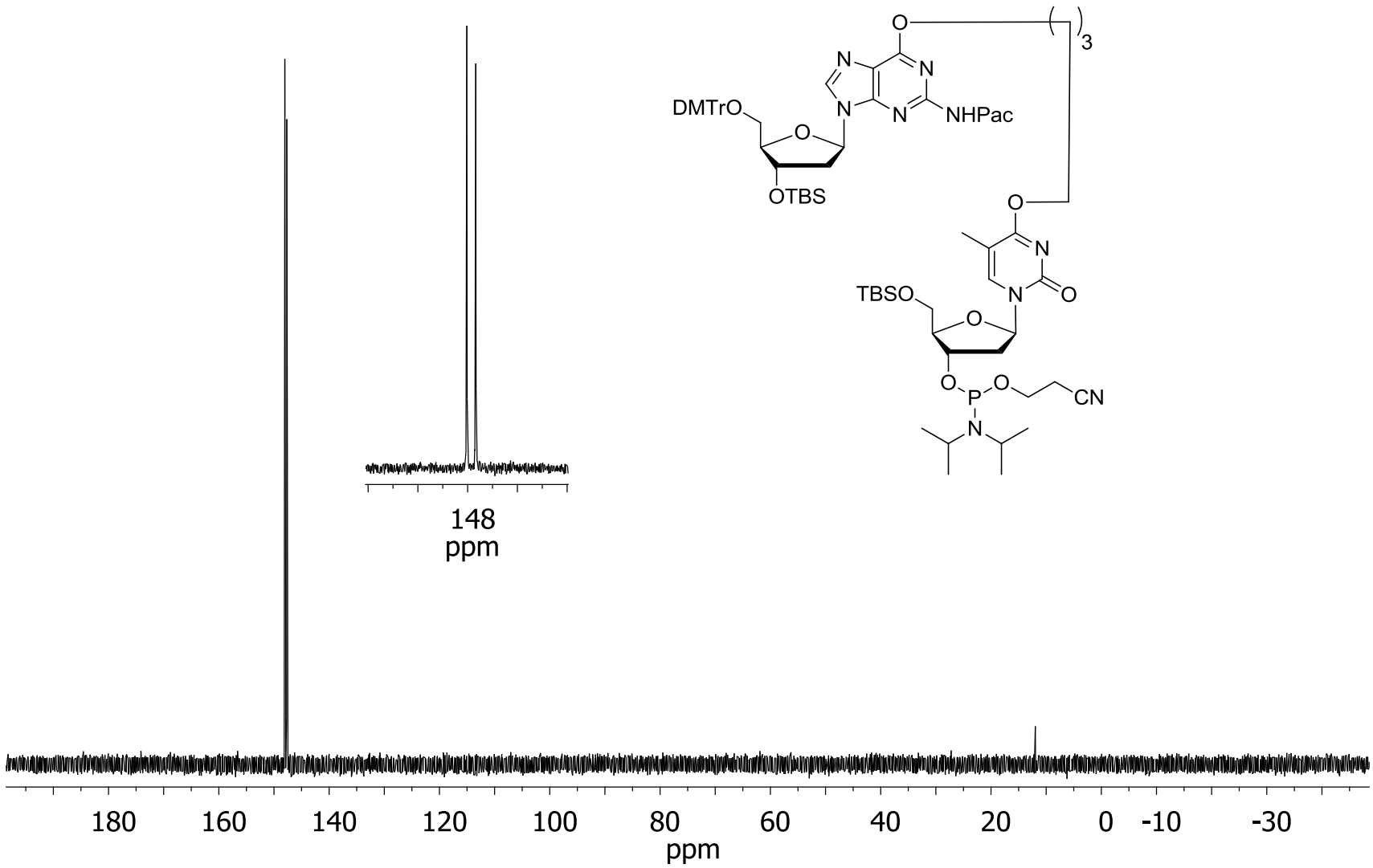
Supplementary Figure 29 - 500 MHz ^{1}H NMR spectrum of compound (**11a**) (in d_6-acetone)



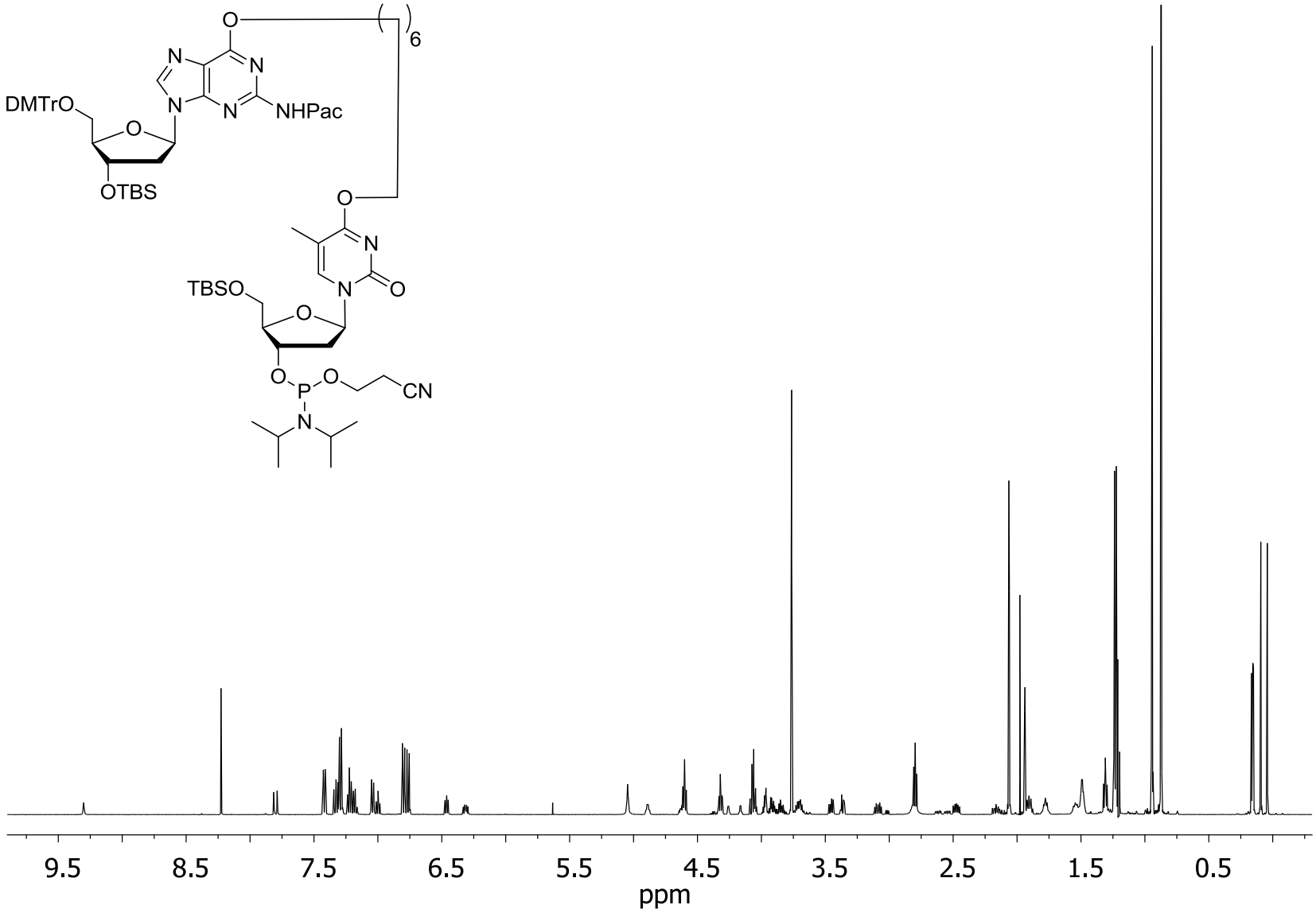
Supplementary Figure 30 - 125.7 MHz ^{13}C NMR spectrum of compound (**11a**) (in d_6-acetone)



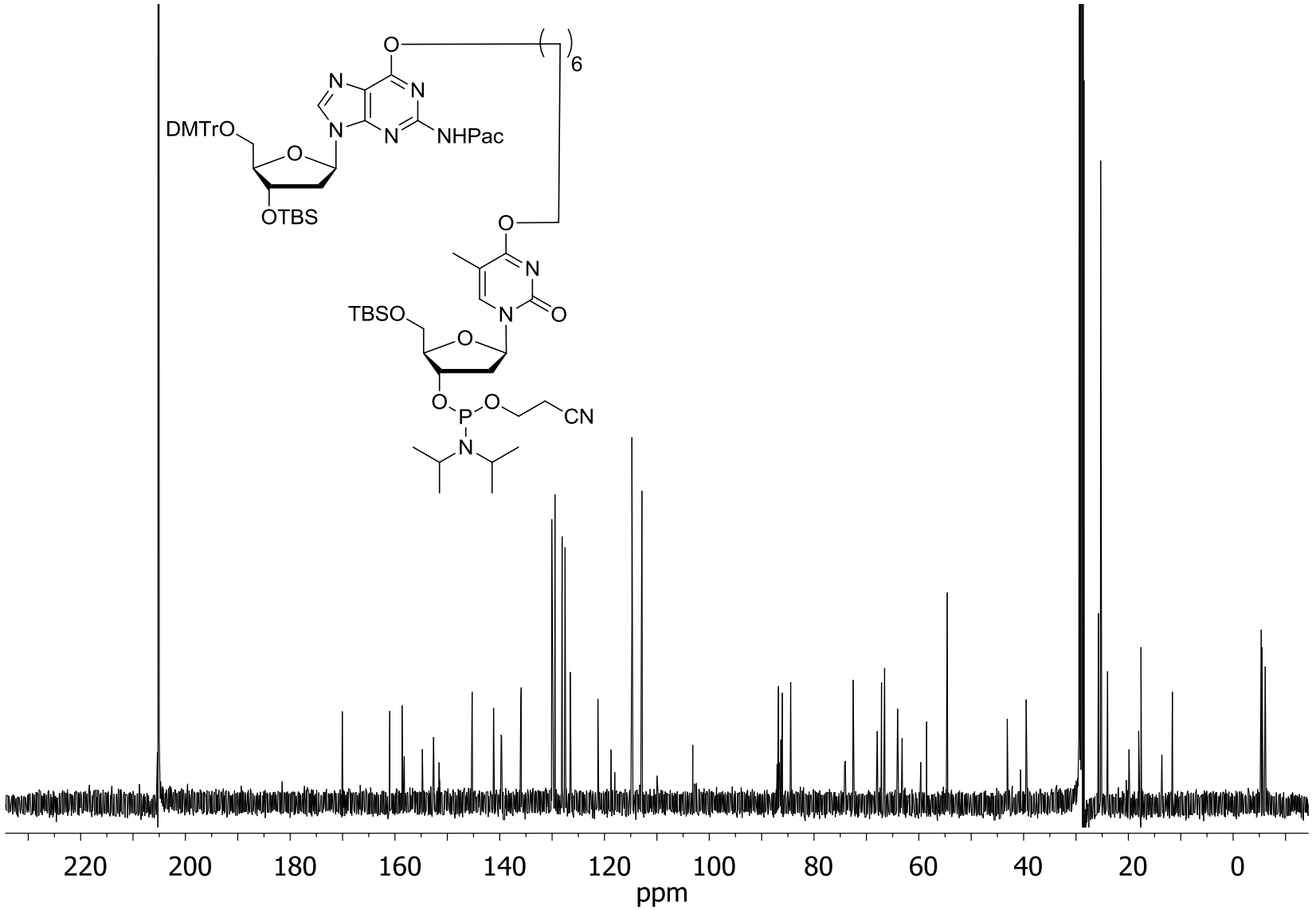
Supplementary Figure 31 - 202.3 MHz ^{31}P NMR spectrum of compound (**11a**) (d_6-acetone)



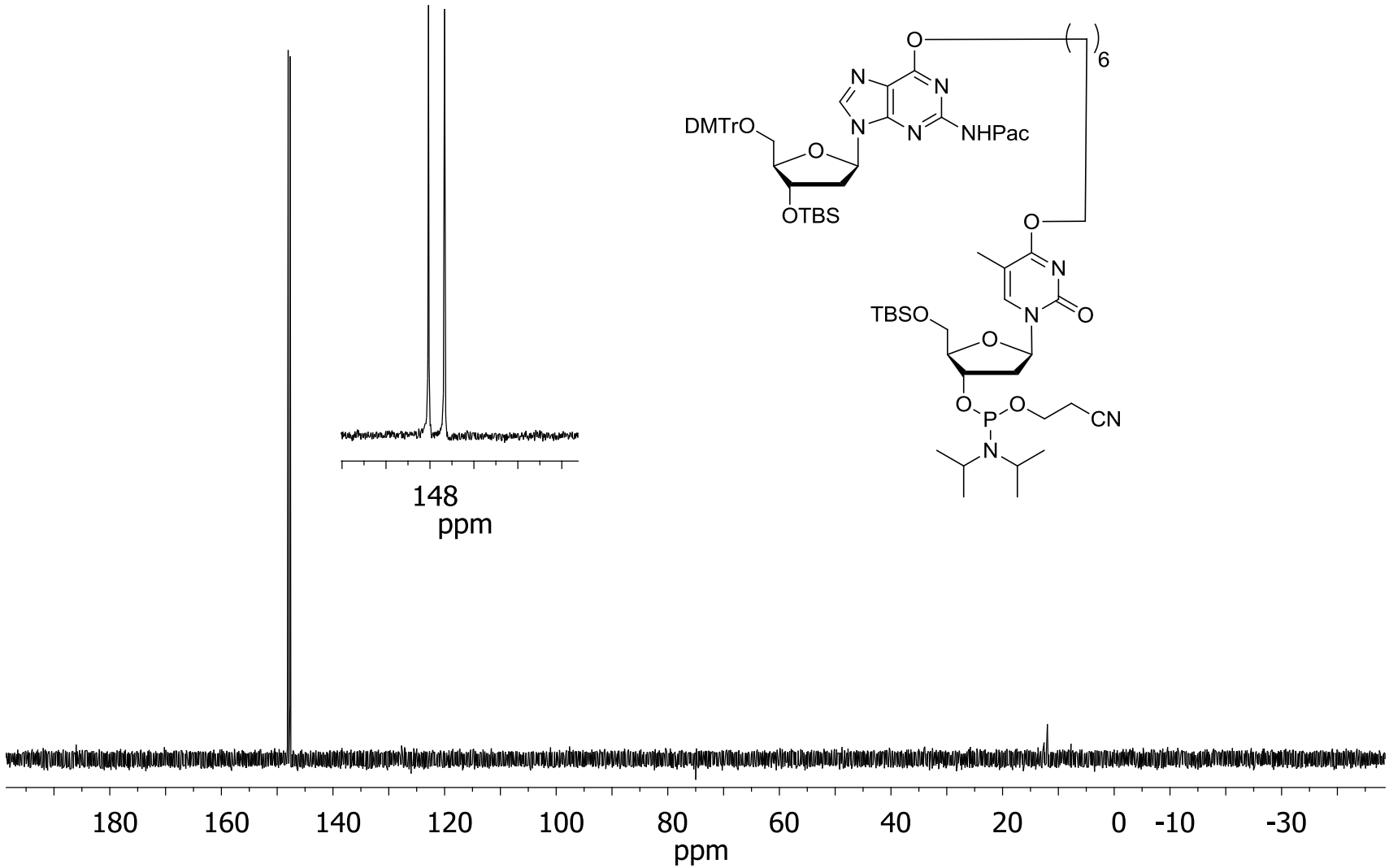
Supplementary Figure 32 - 500 MHz ^{1}H NMR spectrum of compound (**11b**) (in d_6-acetone)



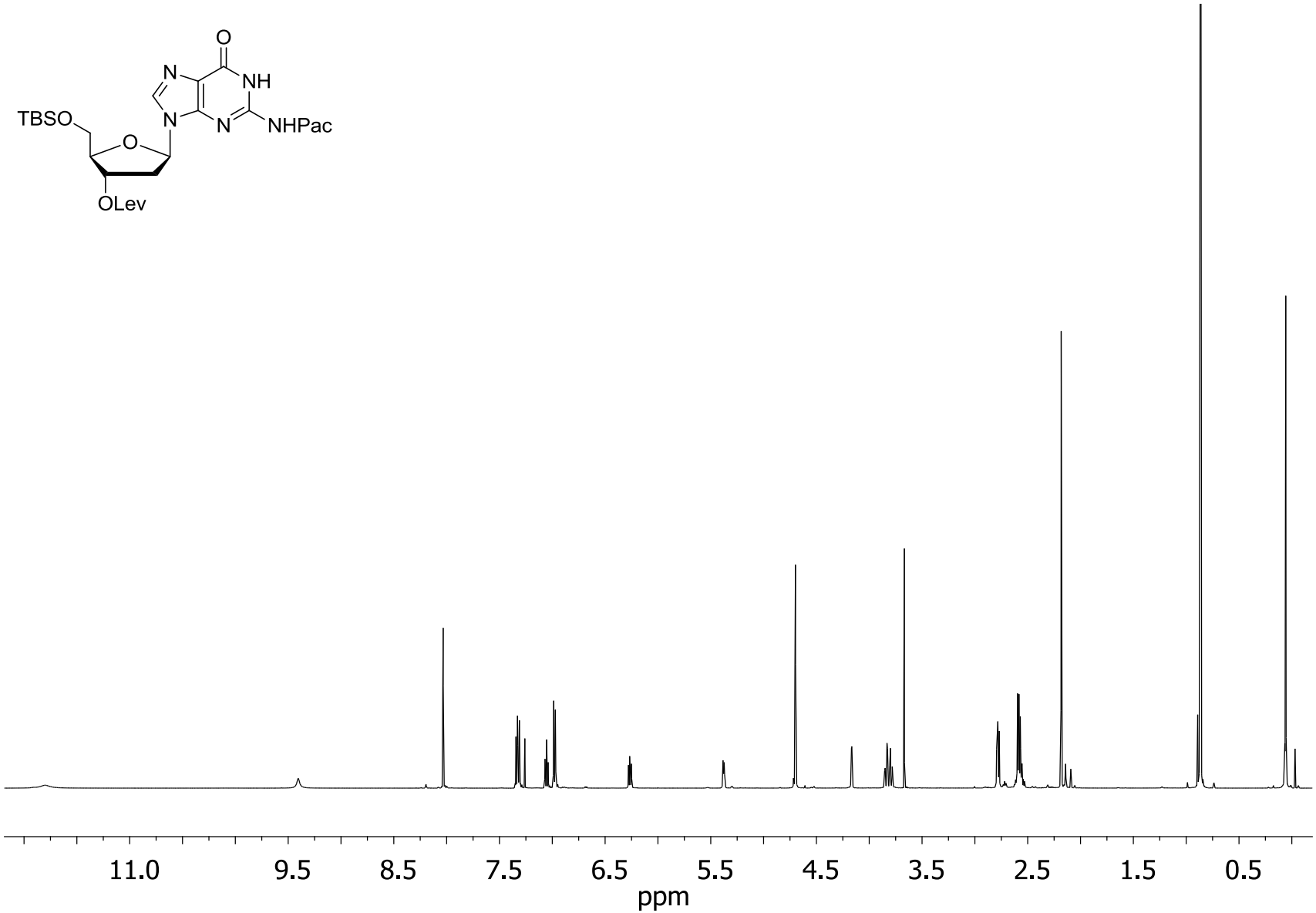
Supplementary Figure 33 - 125.7 MHz ^{13}C NMR spectrum of compound (**11b**) (in d_6-acetone)



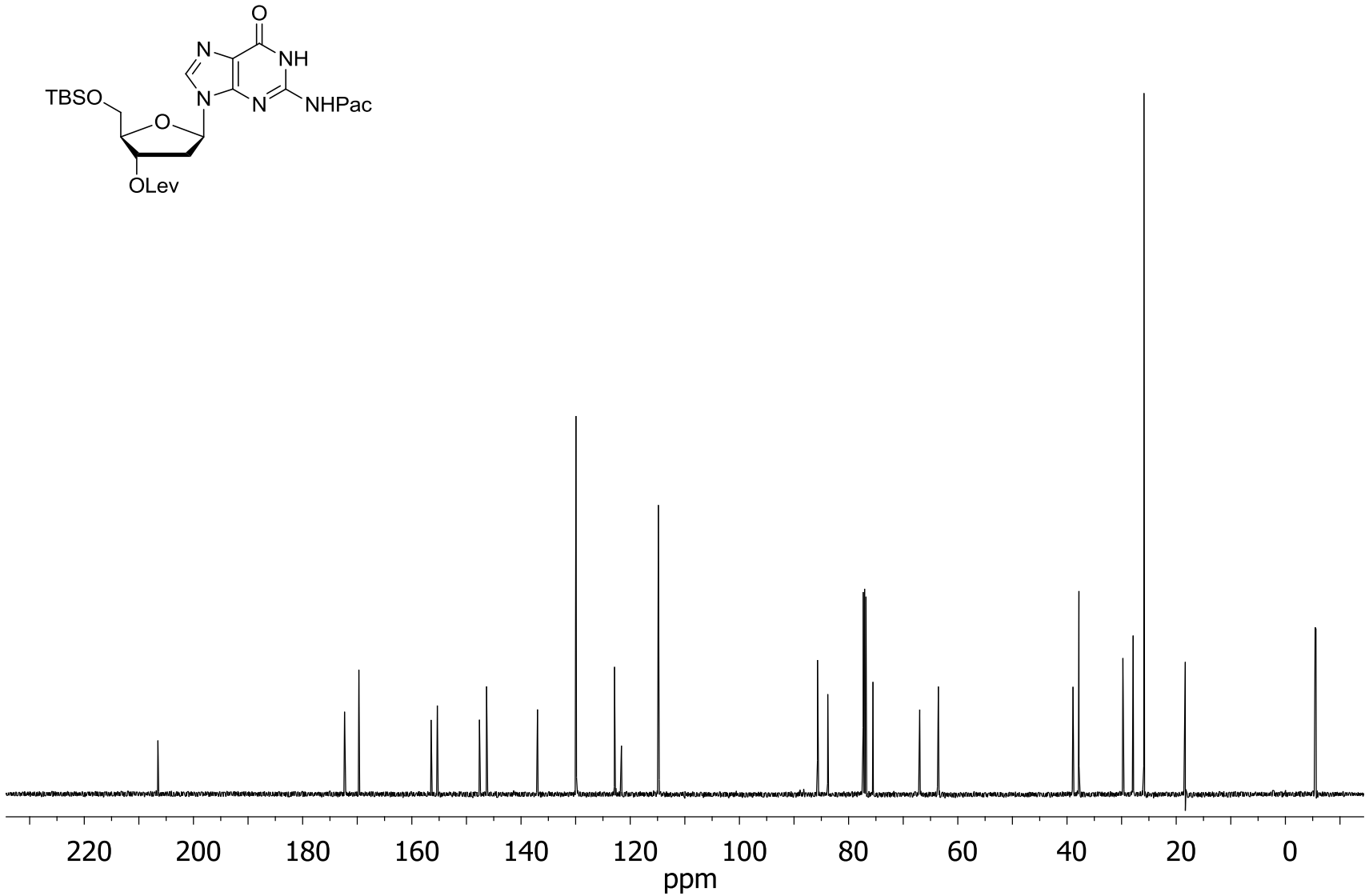
Supplementary Figure 34 - 202.3 MHz ^{31}P NMR spectrum of compound (**11b**) (in d_6-acetone)



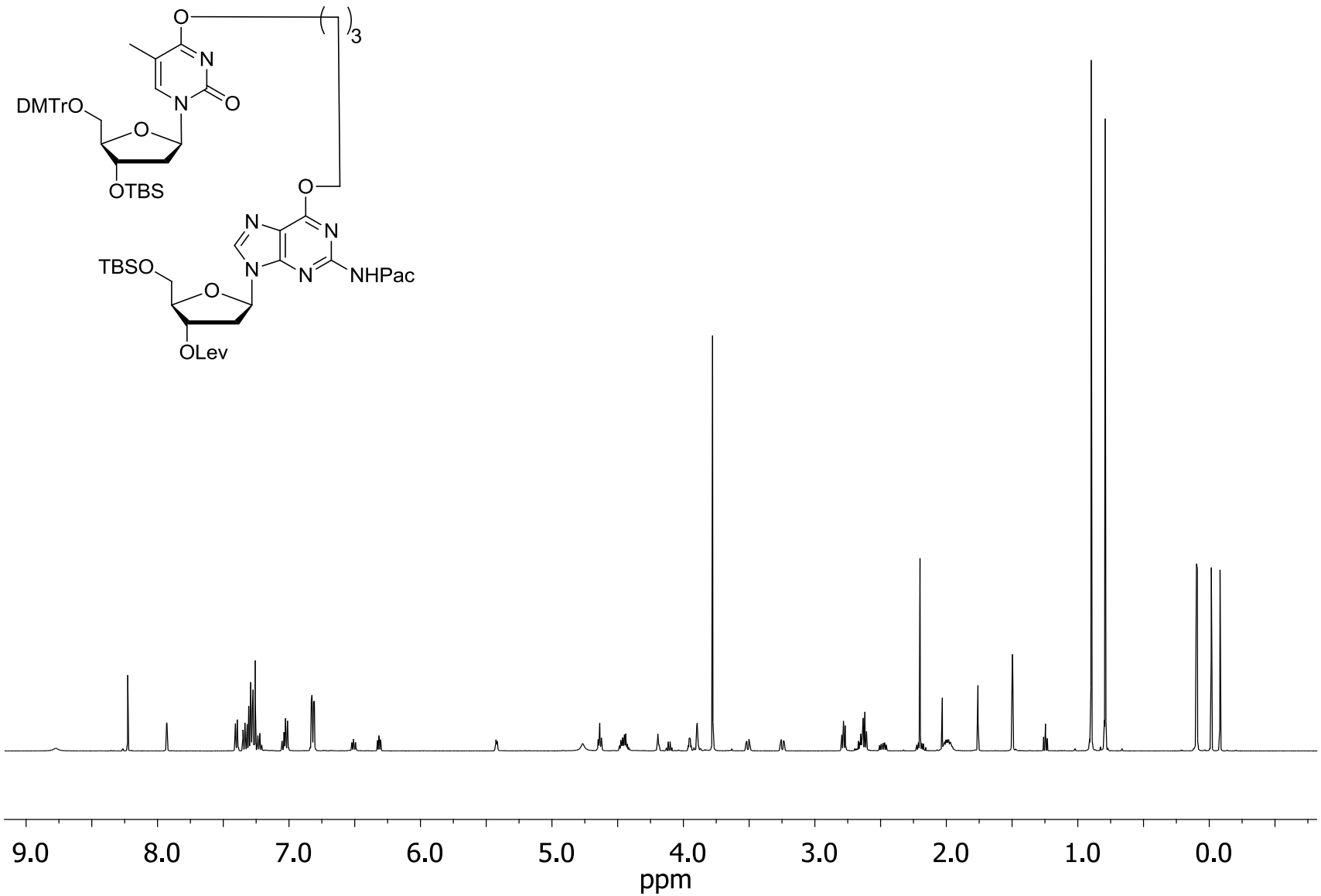
Supplementary Figure 35 - 500 MHz ^{1}H NMR spectrum of compound (**13**) (in $CDCl_3$)



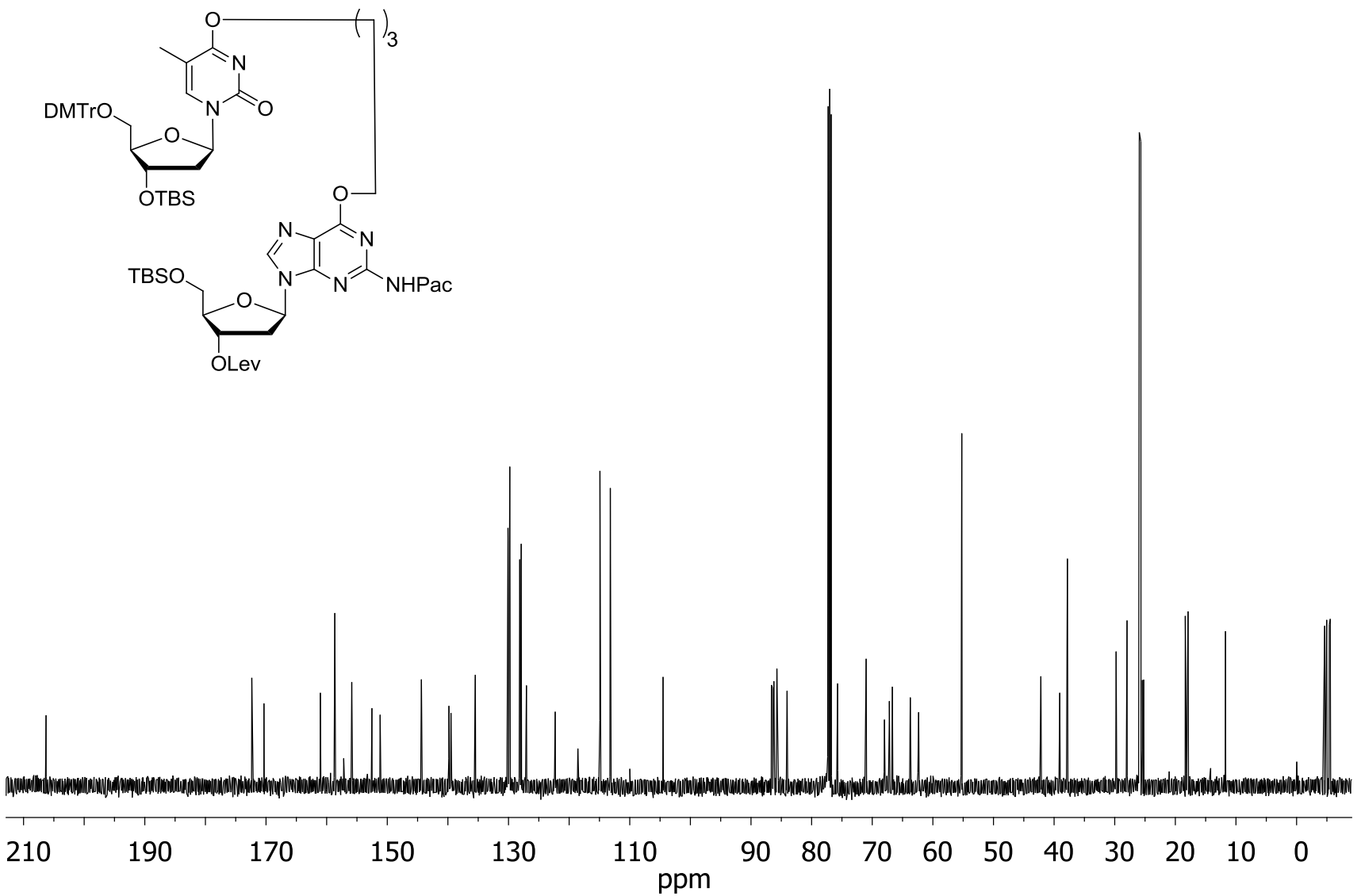
Supplementary Figure 36 - 125.7 MHz ^{13}C NMR spectrum of compound (**13**) (in $CDCl_3$)



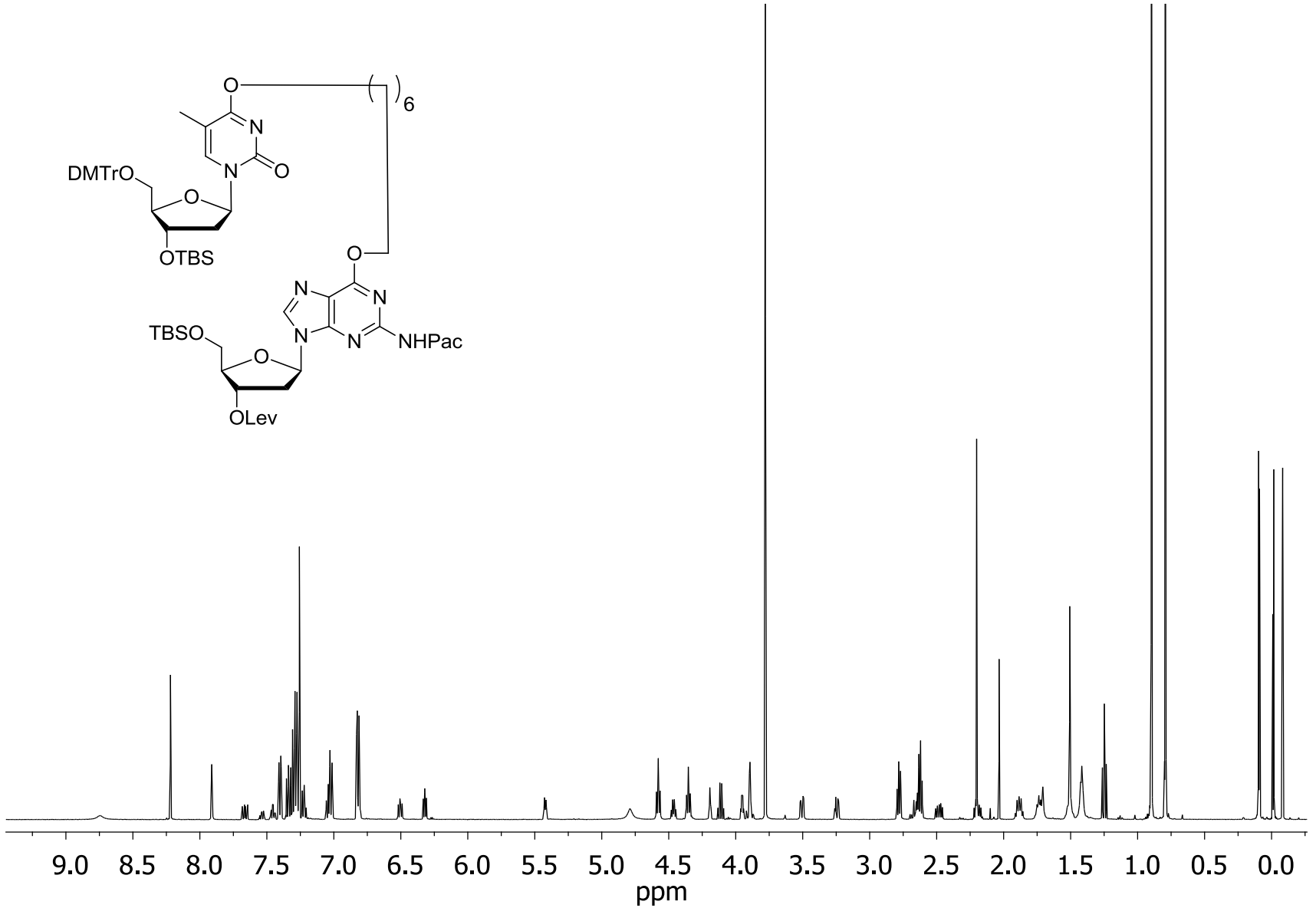
Supplementary Figure 37 - 500 MHz ^{1}H NMR spectrum of compound (**14a**) (in $CDCl_3$)



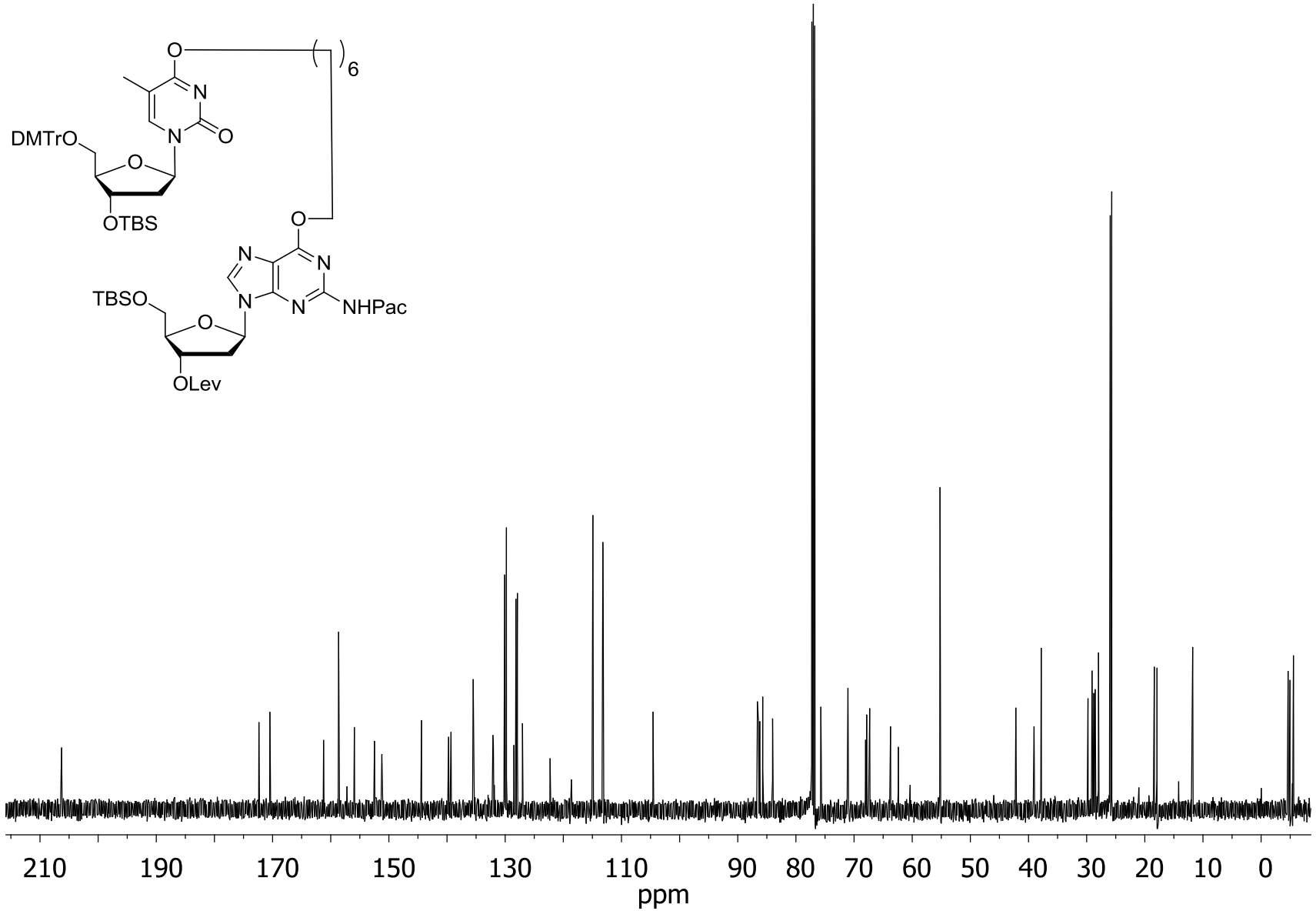
Supplementary Figure 38 - 125.7 MHz ^{13}C NMR spectrum of compound (**14a**) (in $CDCl_3$)



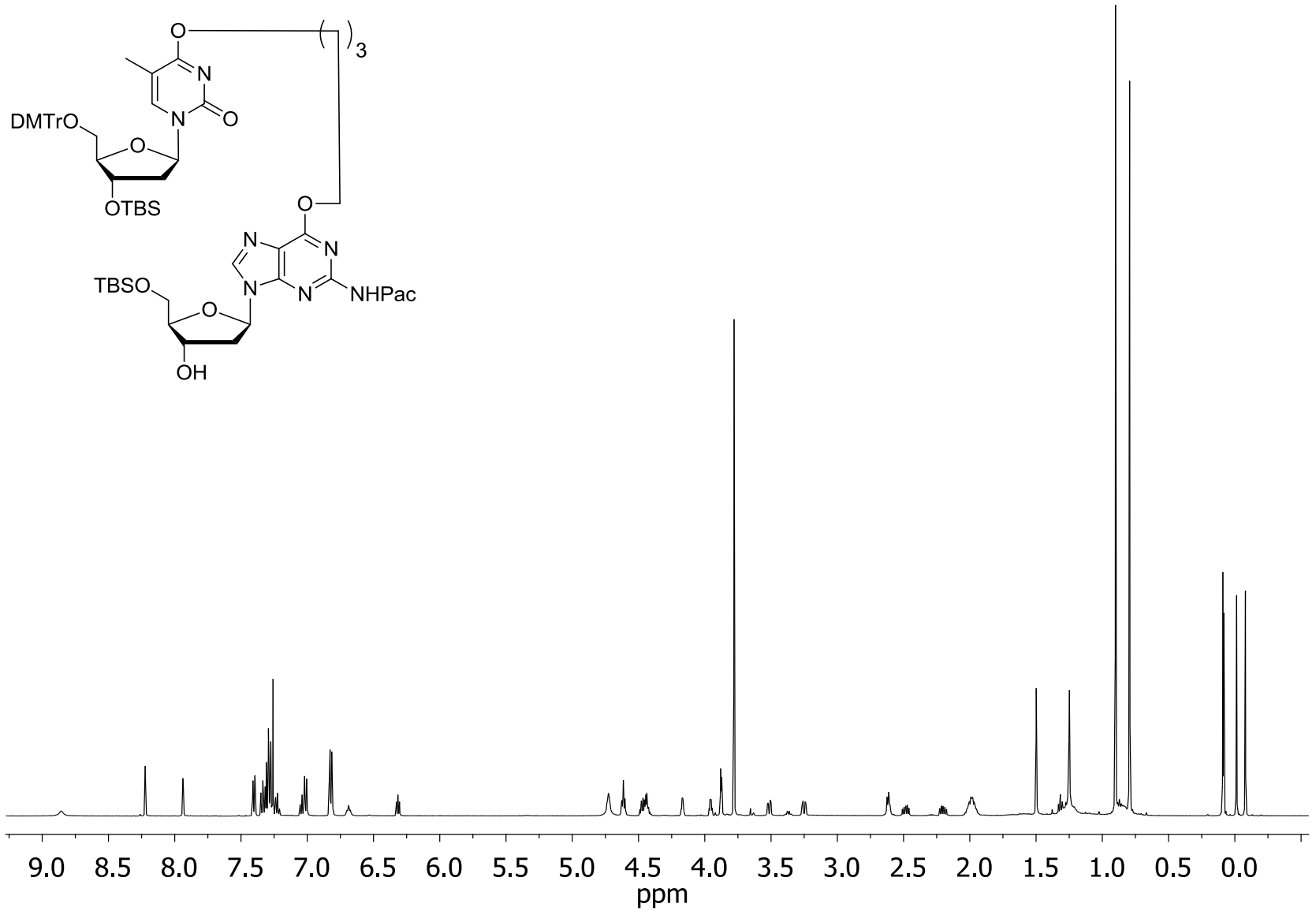
Supplementary Figure 39 - 500 MHz ^{1}H NMR spectrum of compound (**14b**) (in $CDCl_3$)



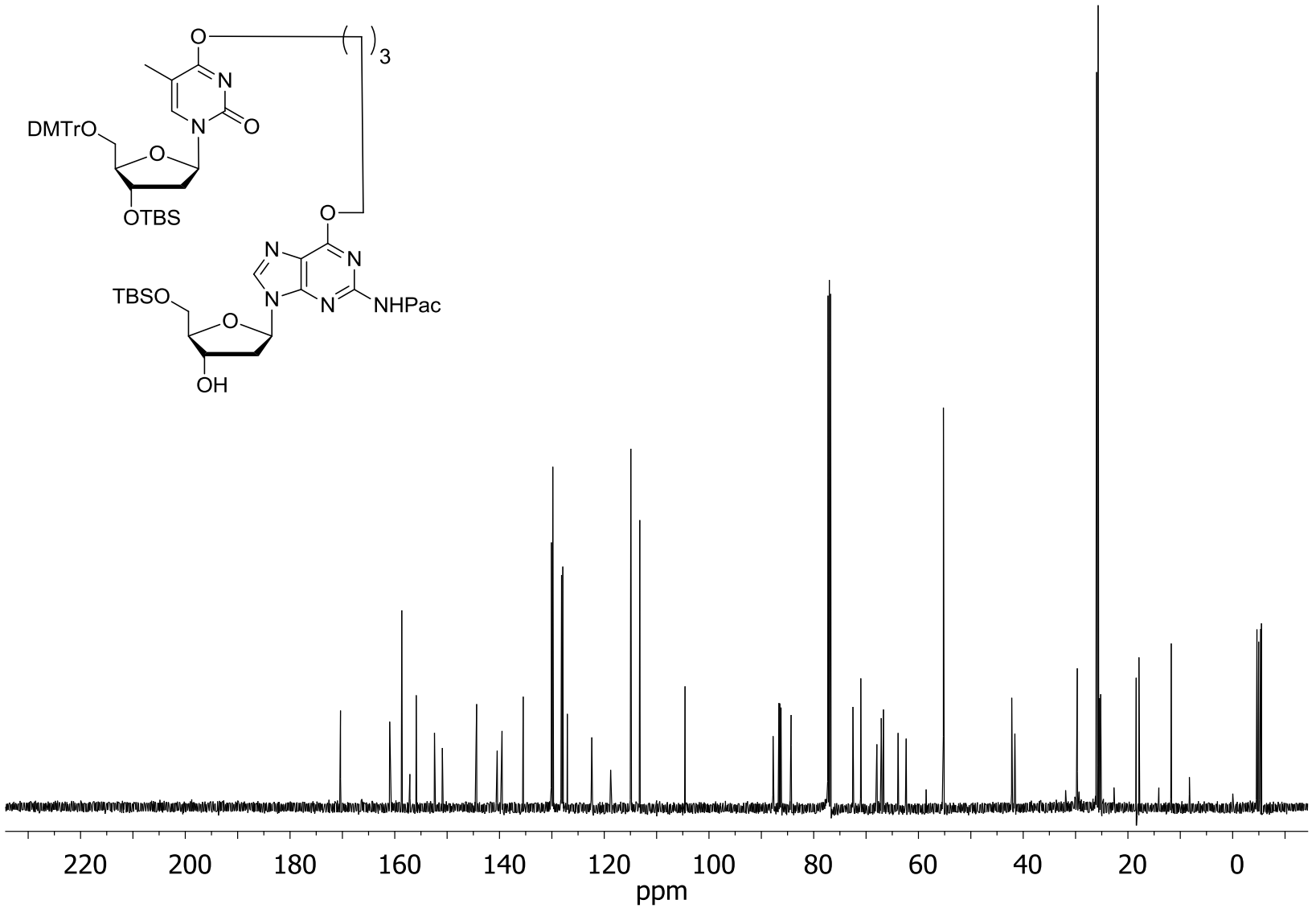
Supplementary Figure 40 - 125.7 MHz ^{13}C NMR spectrum of compound (**14b**) (in $CDCl_3$)



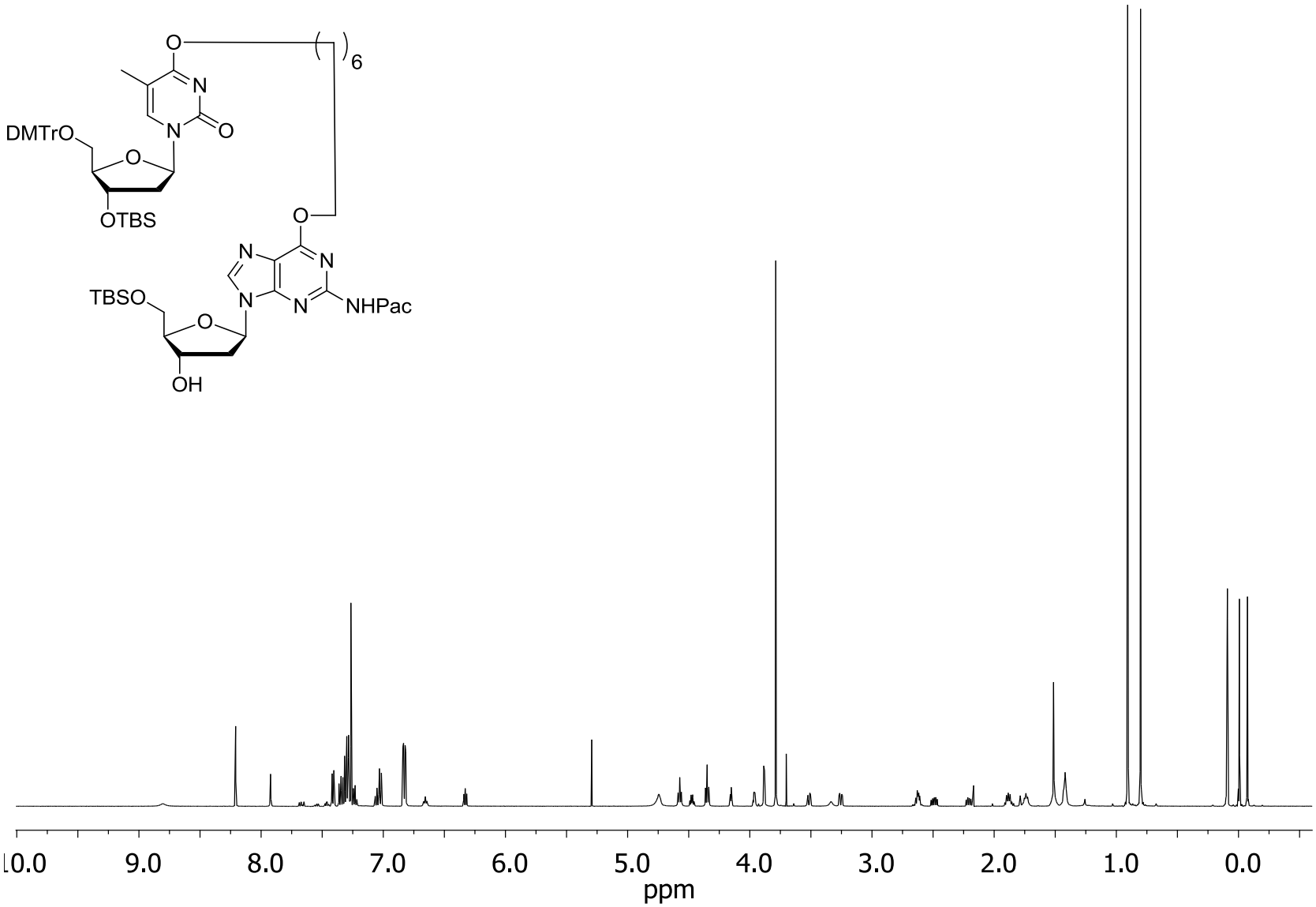
Supplementary Figure 41 - 500 MHz ^{1}H NMR spectrum of compound (**15a**) (in $CDCl_3$)



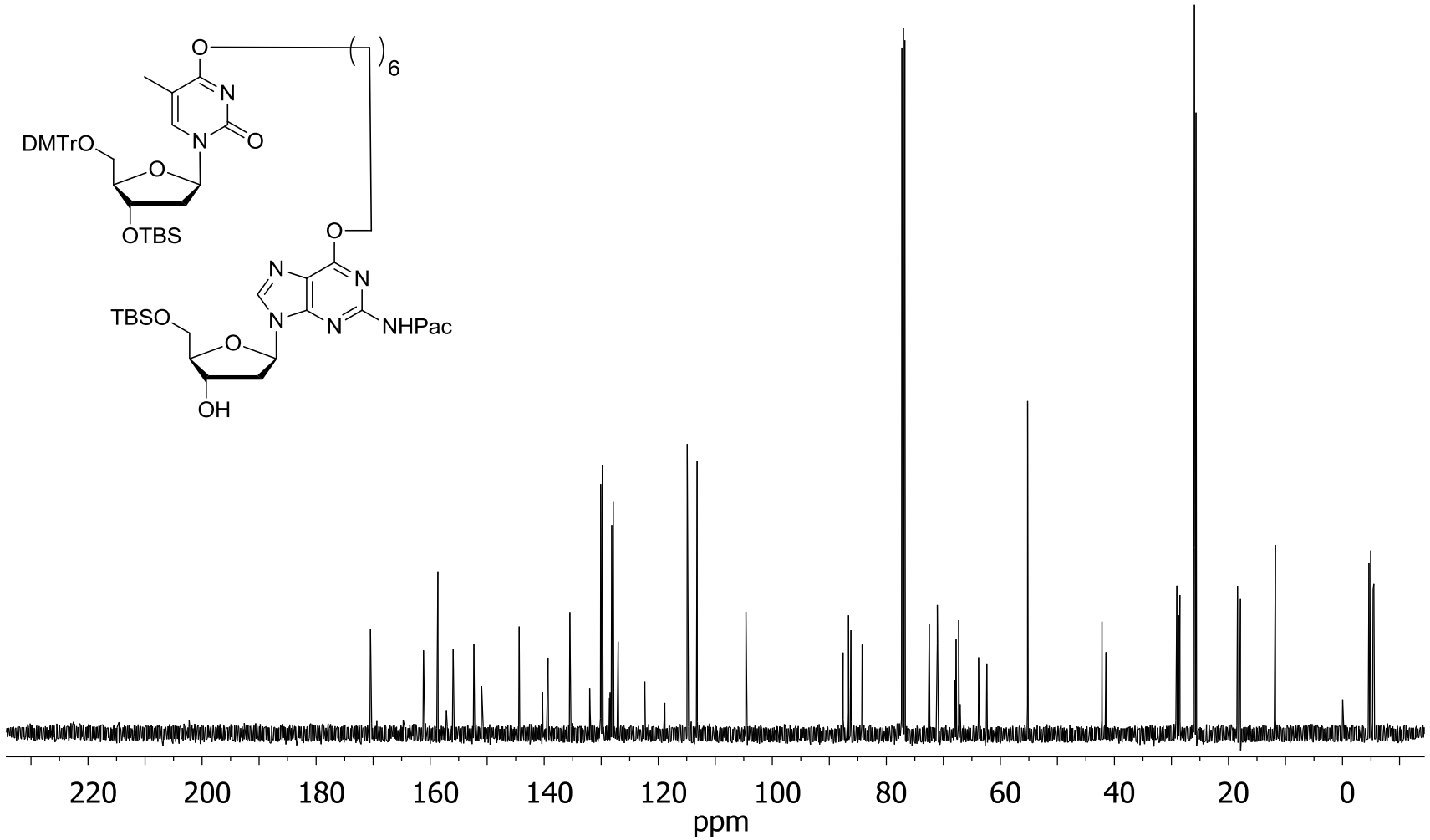
Supplementary Figure 42 - 125.7 MHz ^{13}C NMR spectrum of compound (**15a**) (in $CDCl_3$)



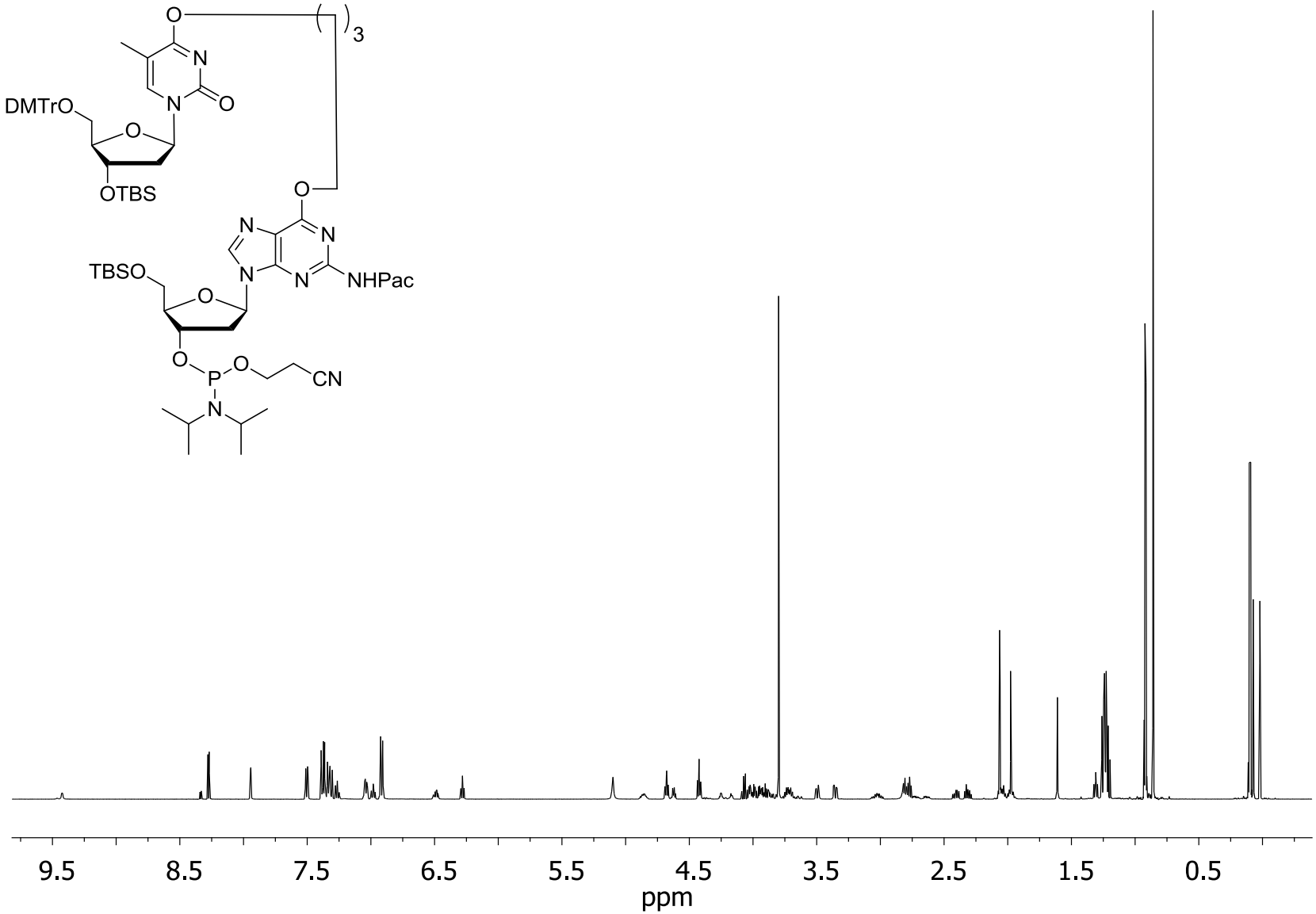
Supplementary Figure 43 - 500 MHz ^{1}H NMR spectrum of compound (**15b**) (in $CDCl_3$)



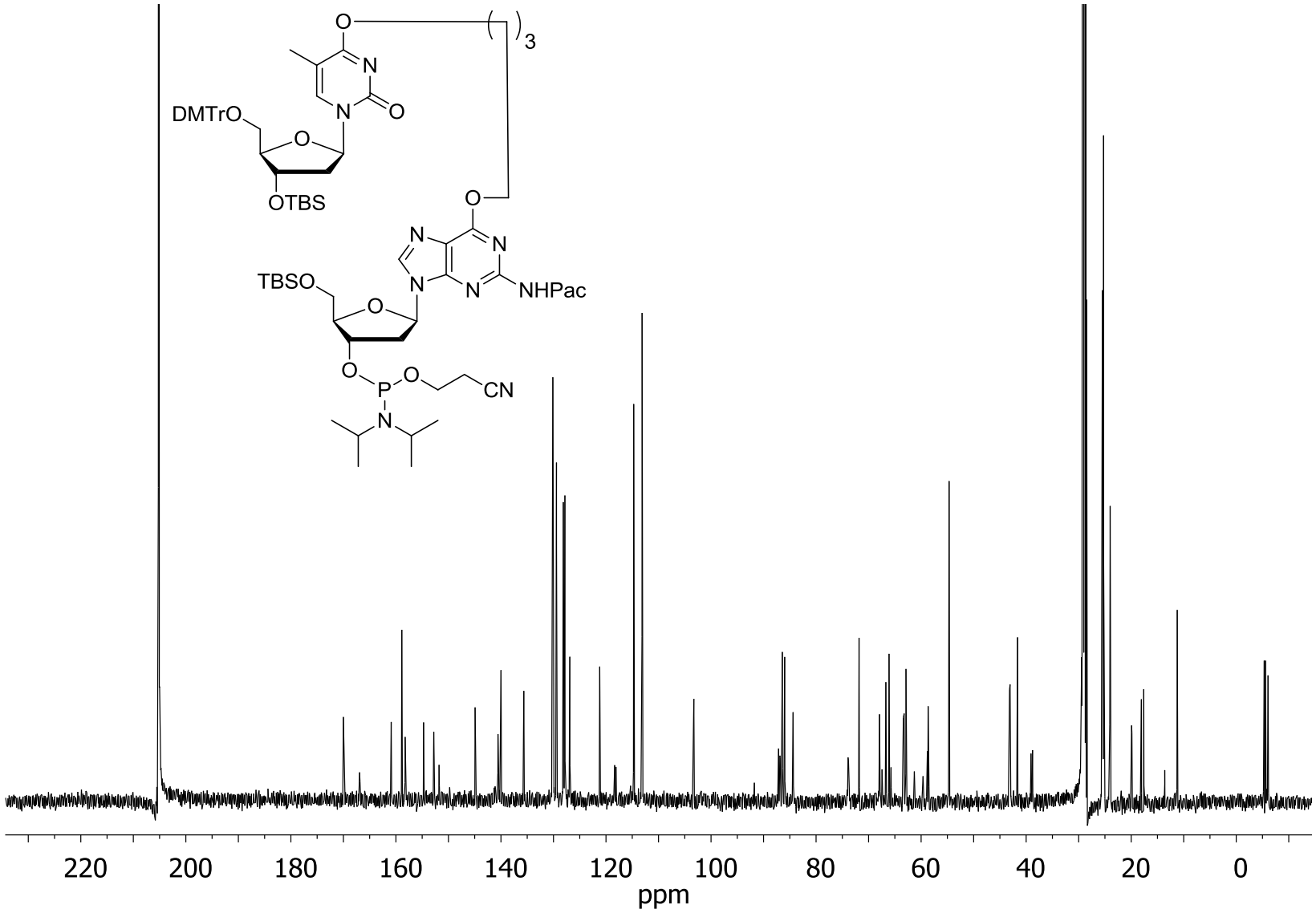
Supplementary Figure 44 - 125.7 MHz ^{13}C NMR spectrum of compound (**15b**) (in $CDCl_3$)



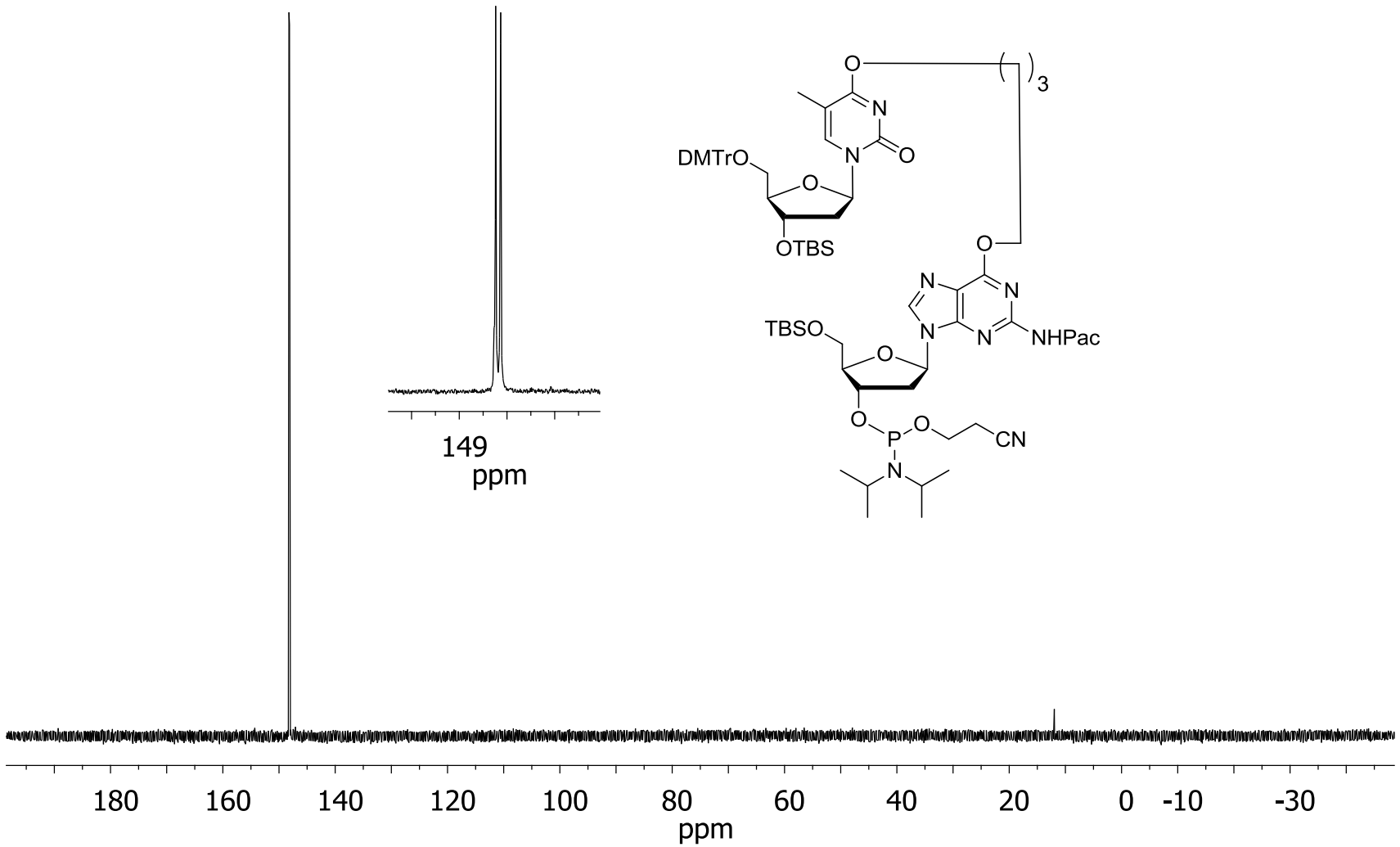
Supplementary Figure 45 - 500 MHz ^{1}H NMR spectrum of compound (**16a**) (in d_6-acetone)



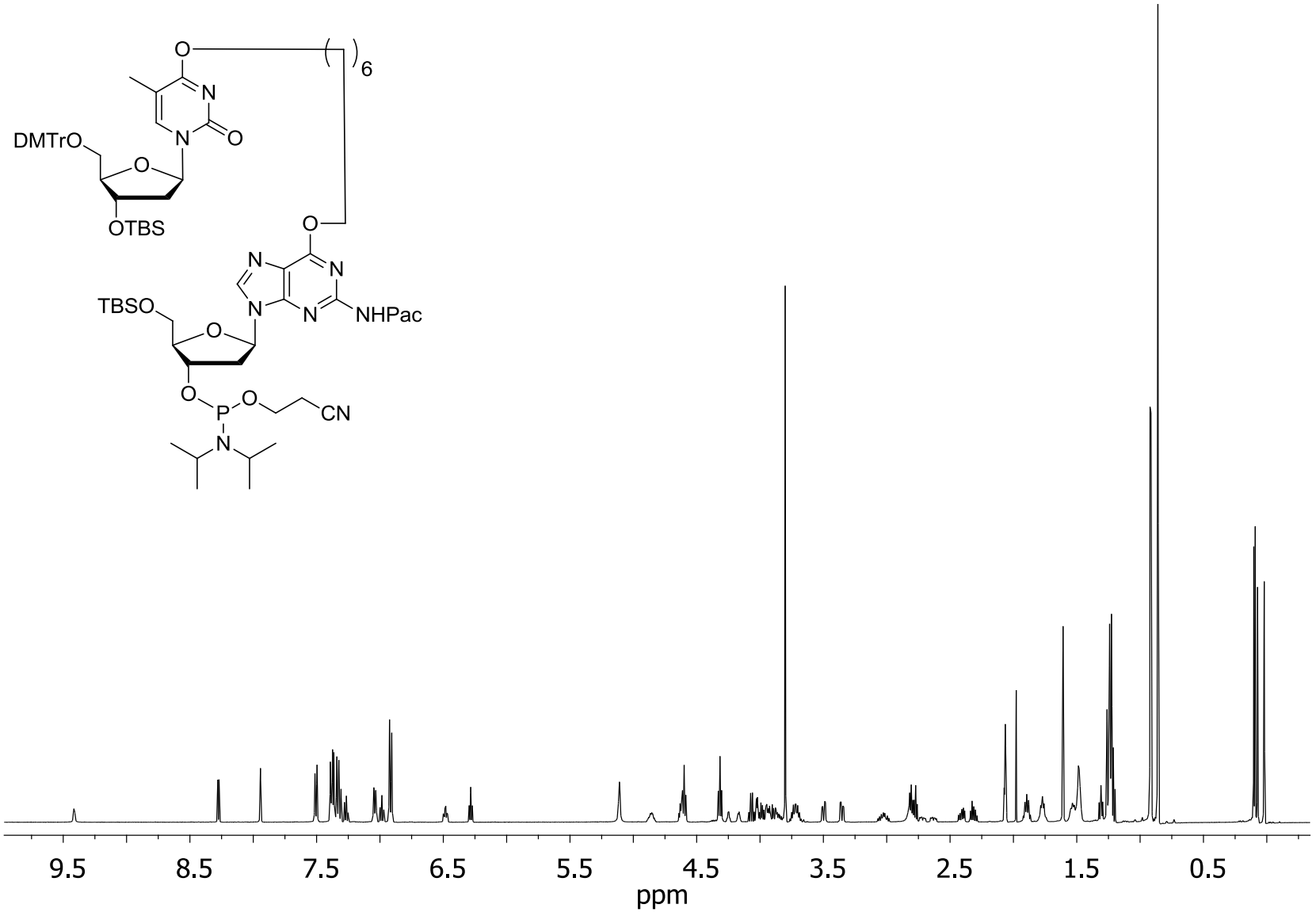
Supplementary Figure 46 - 125.7 MHz ^{13}C NMR spectrum of compound (**16a**) (in d_6-acetone)



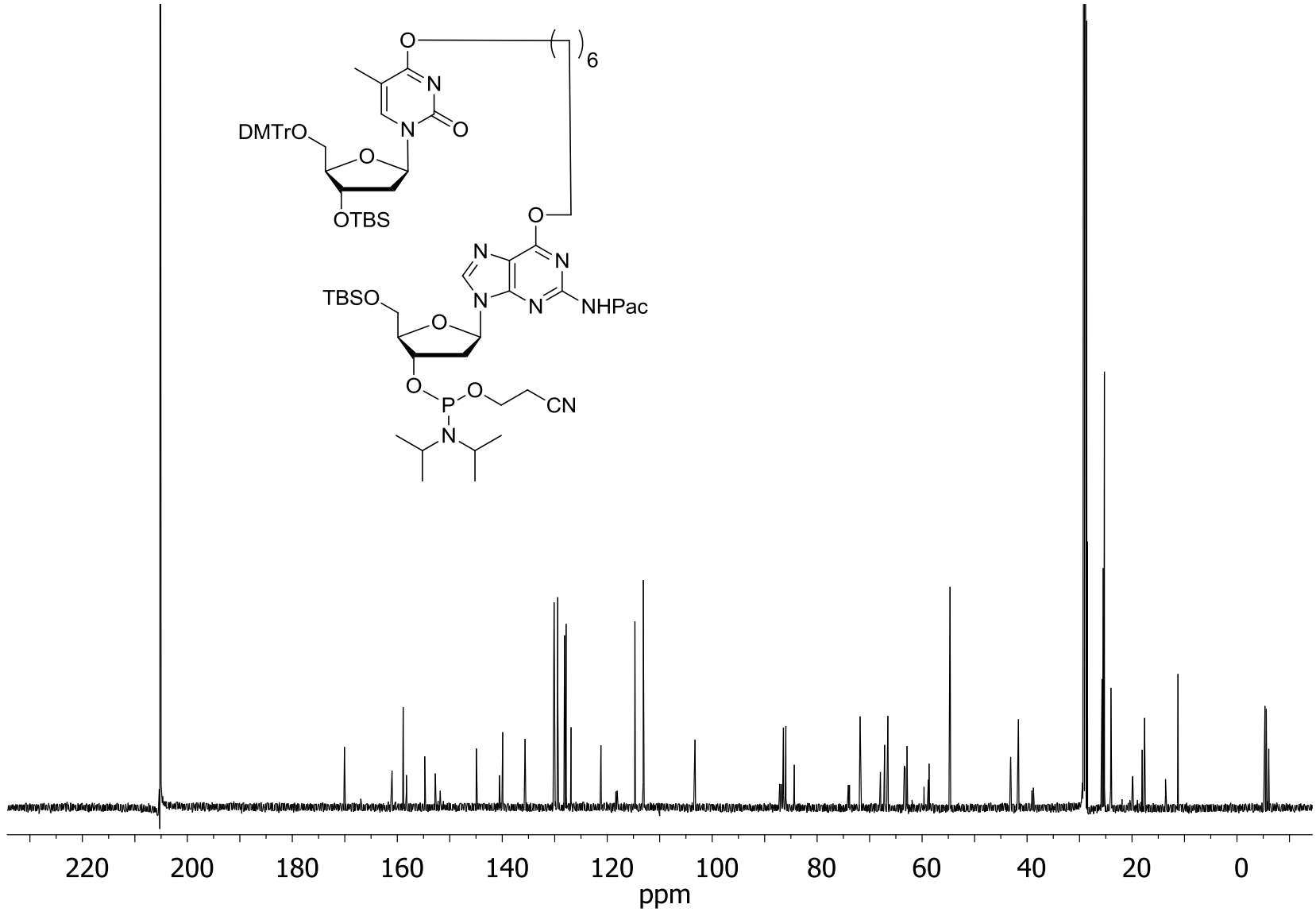
Supplementary Figure 47 - 202.3 MHz ^{31}P NMR spectrum of compound (**16a**) (d_6-acetone)



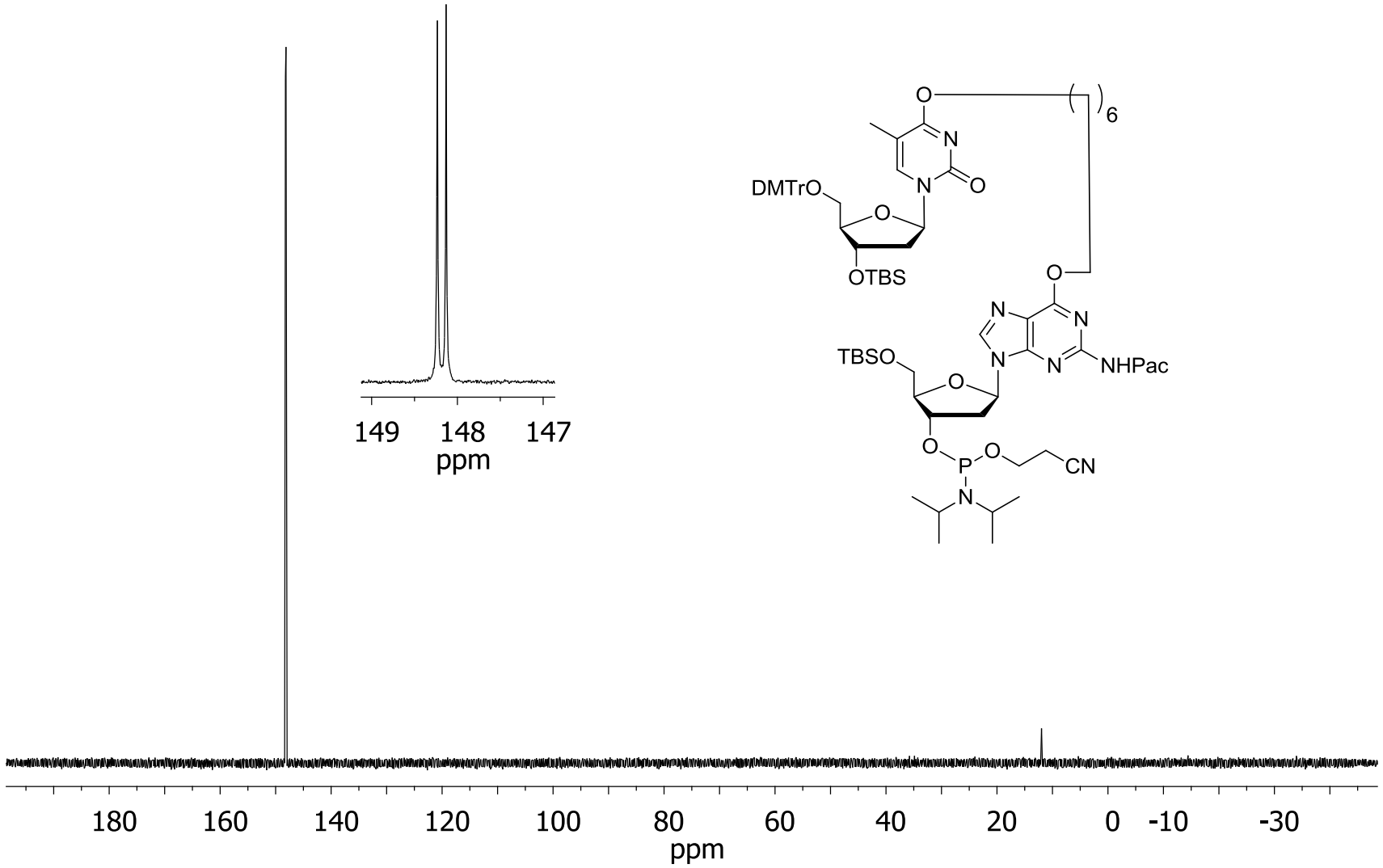
Supplementary Figure 48 - 500 MHz ^{1}H NMR spectrum of compound (**16b**) (in d_6-acetone)



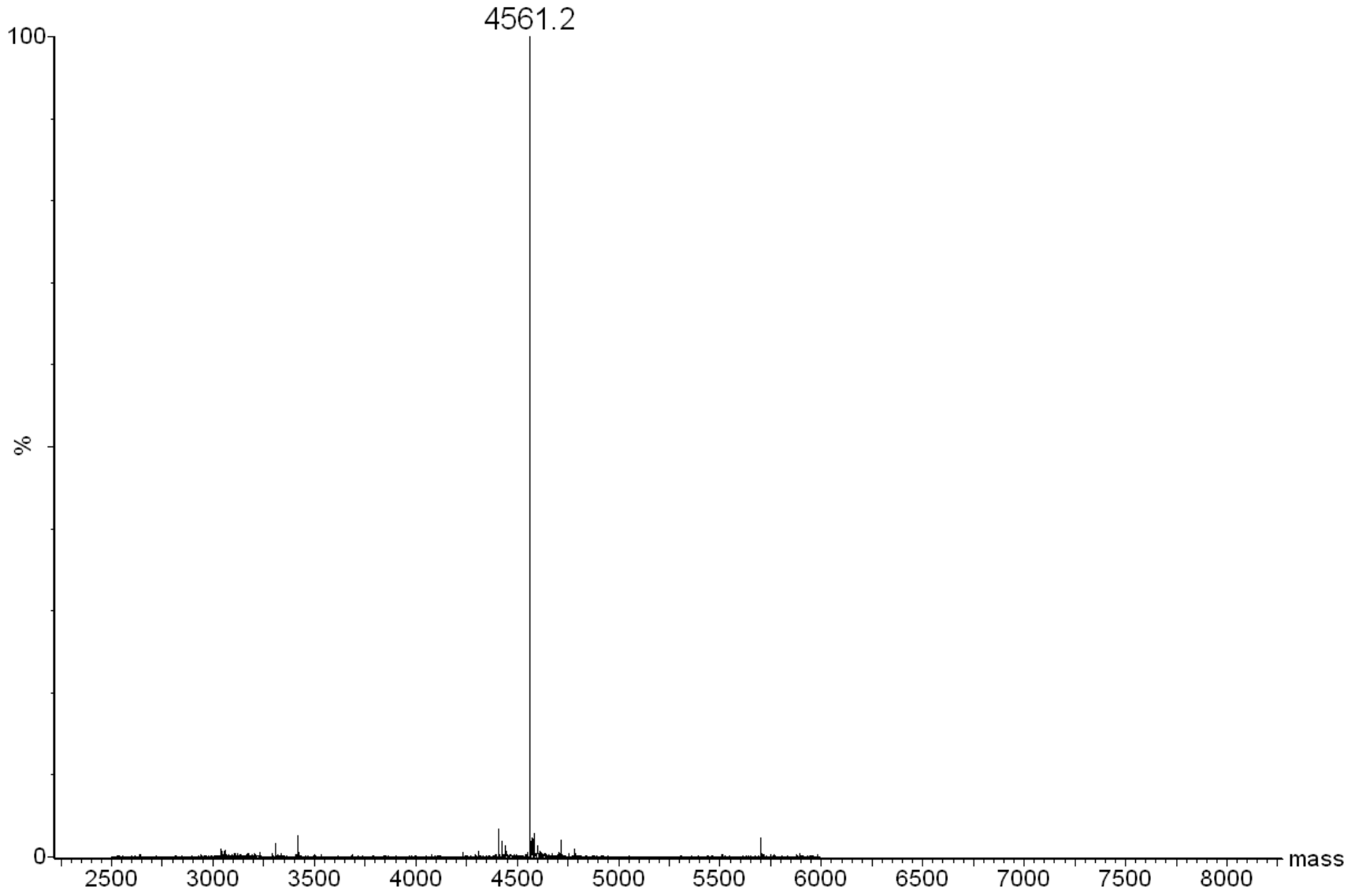
Supplementary Figure 49 - 125.7 MHz ^{13}C NMR spectrum of compound (**16b**) (in d_6-acetone)



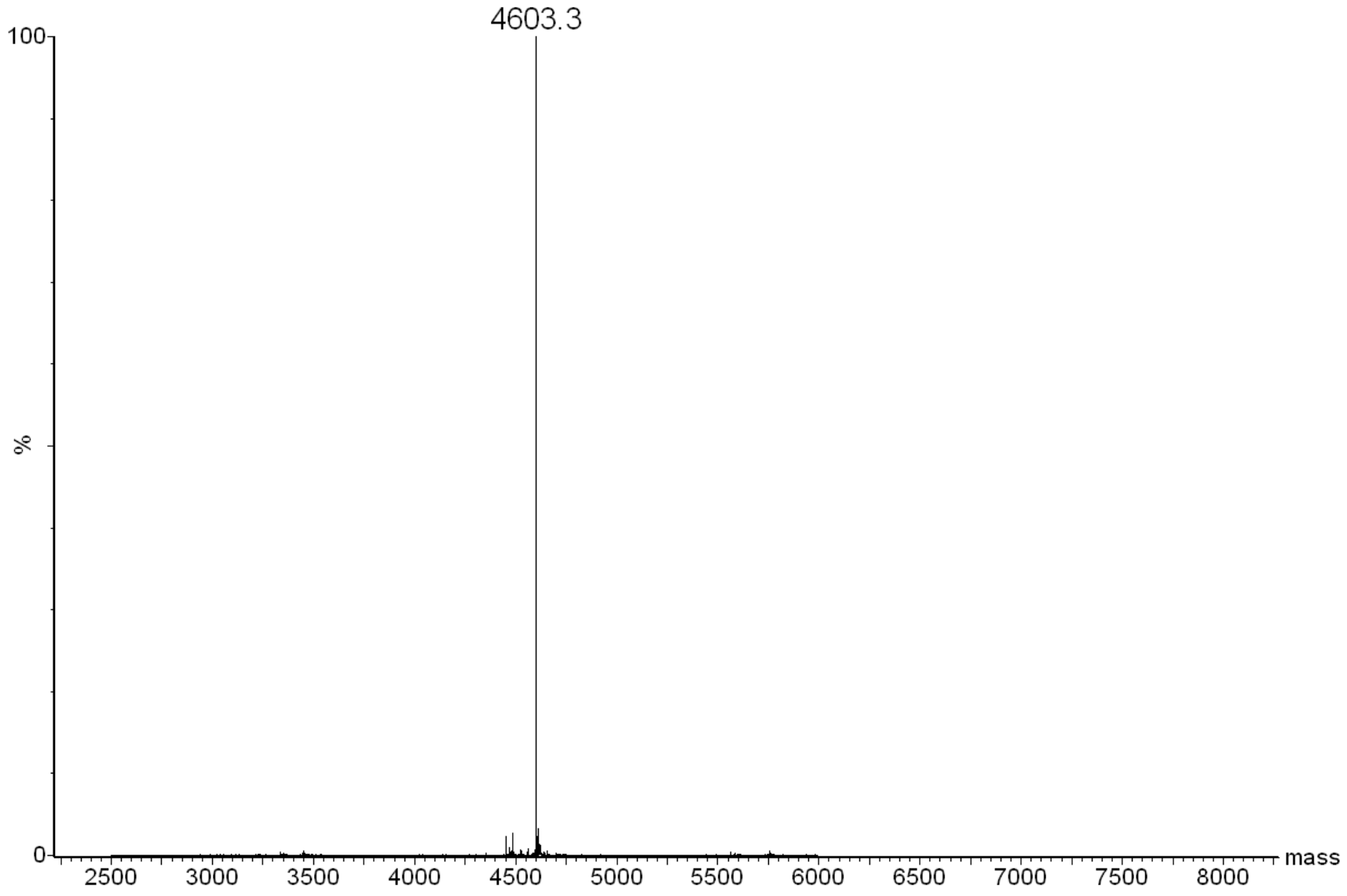
Supplementary Figure 50 - 202.3 MHz ^{31}P NMR spectrum of compound (**16b**) (in d_6-acetone)



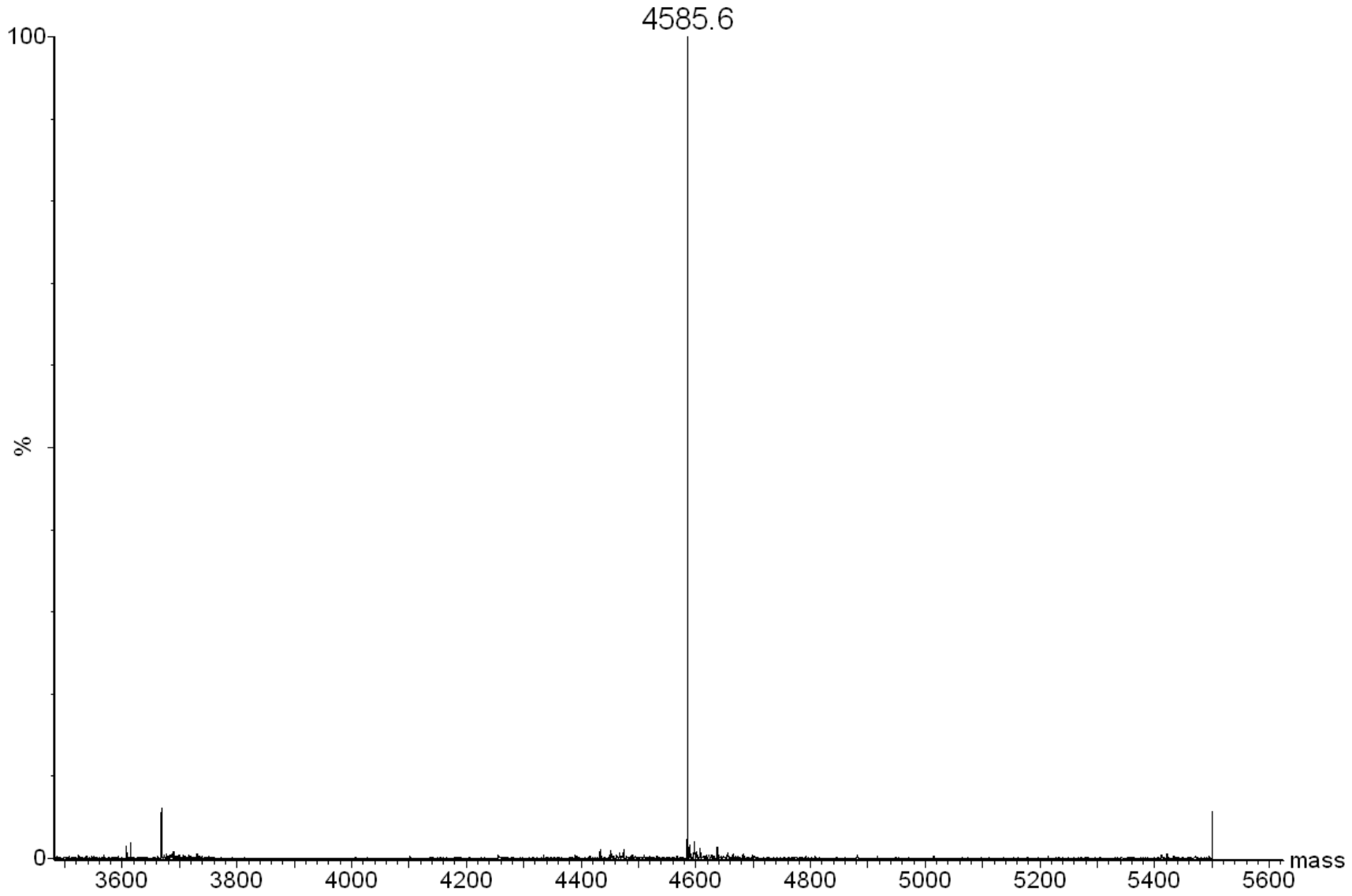
Supplementary Figure 51 - ESI MS spectrum of oligonucleotide **TT4** (expected mass of 4560.1)



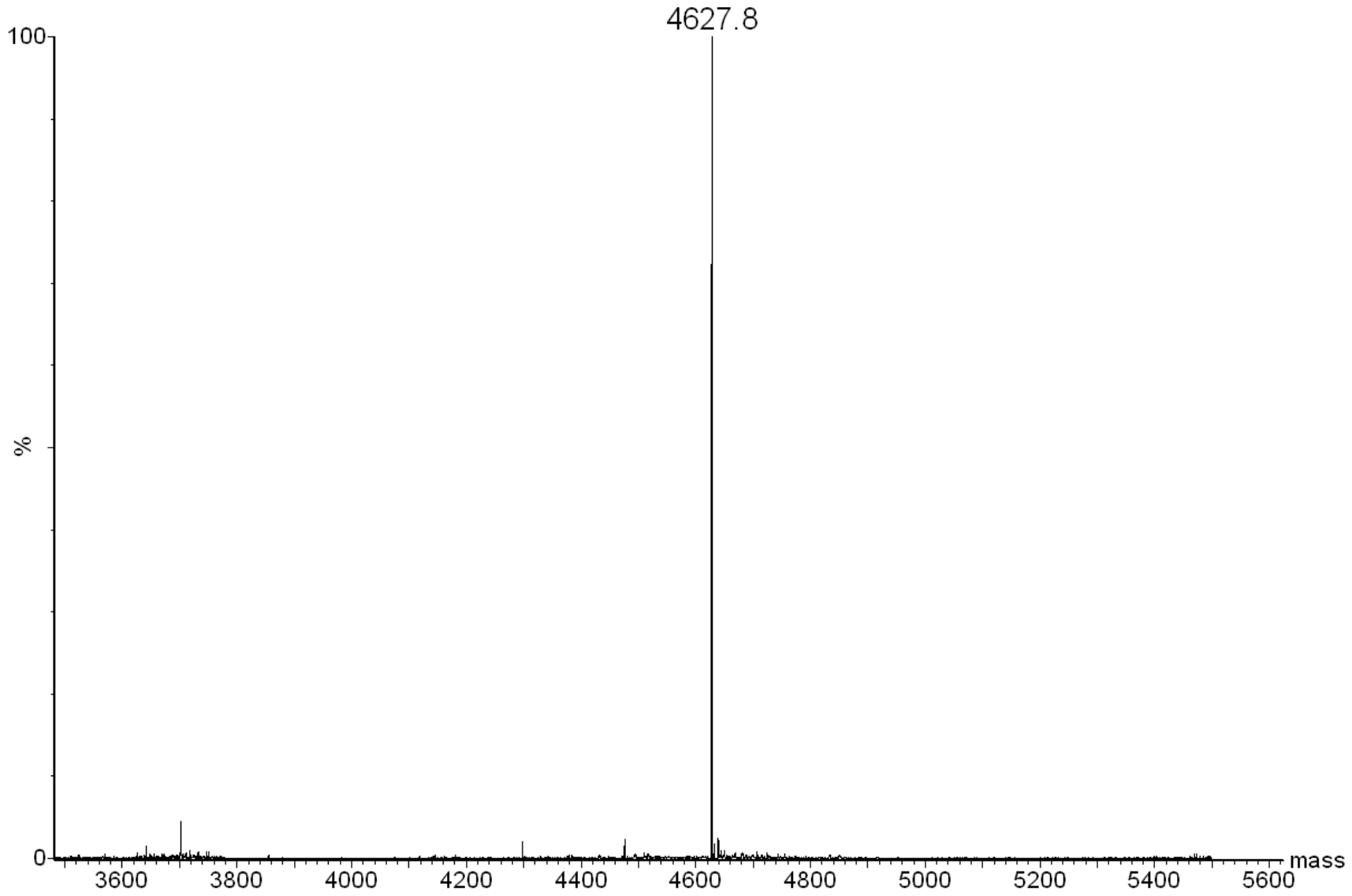
Supplementary Figure 52 - ESI MS spectrum of oligonucleotide **TT7** (expected mass of 4602.2)



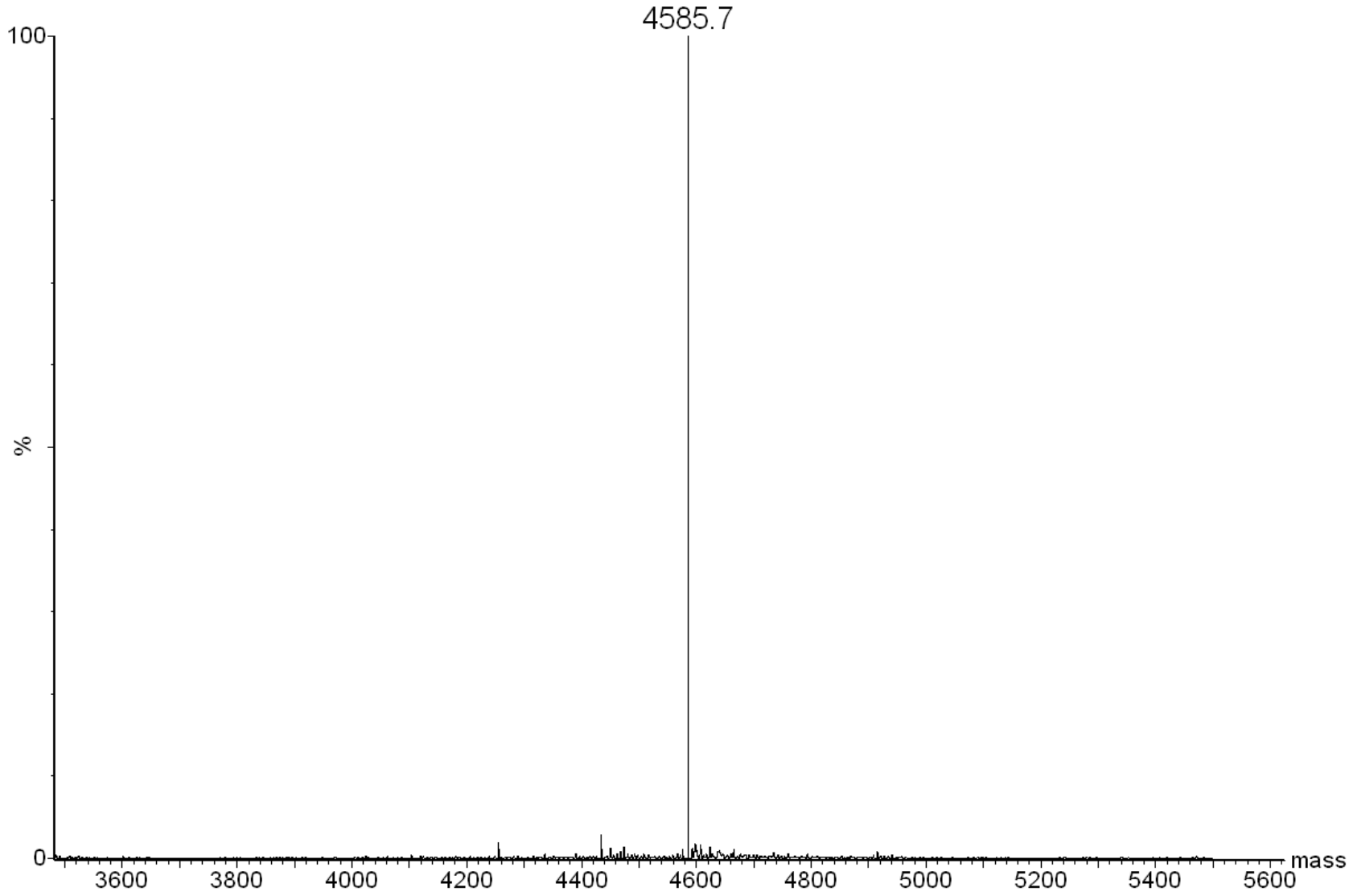
Supplementary Figure 53 - ESI MS spectrum of oligonucleotide **GT4** (expected mass of 4585.1)



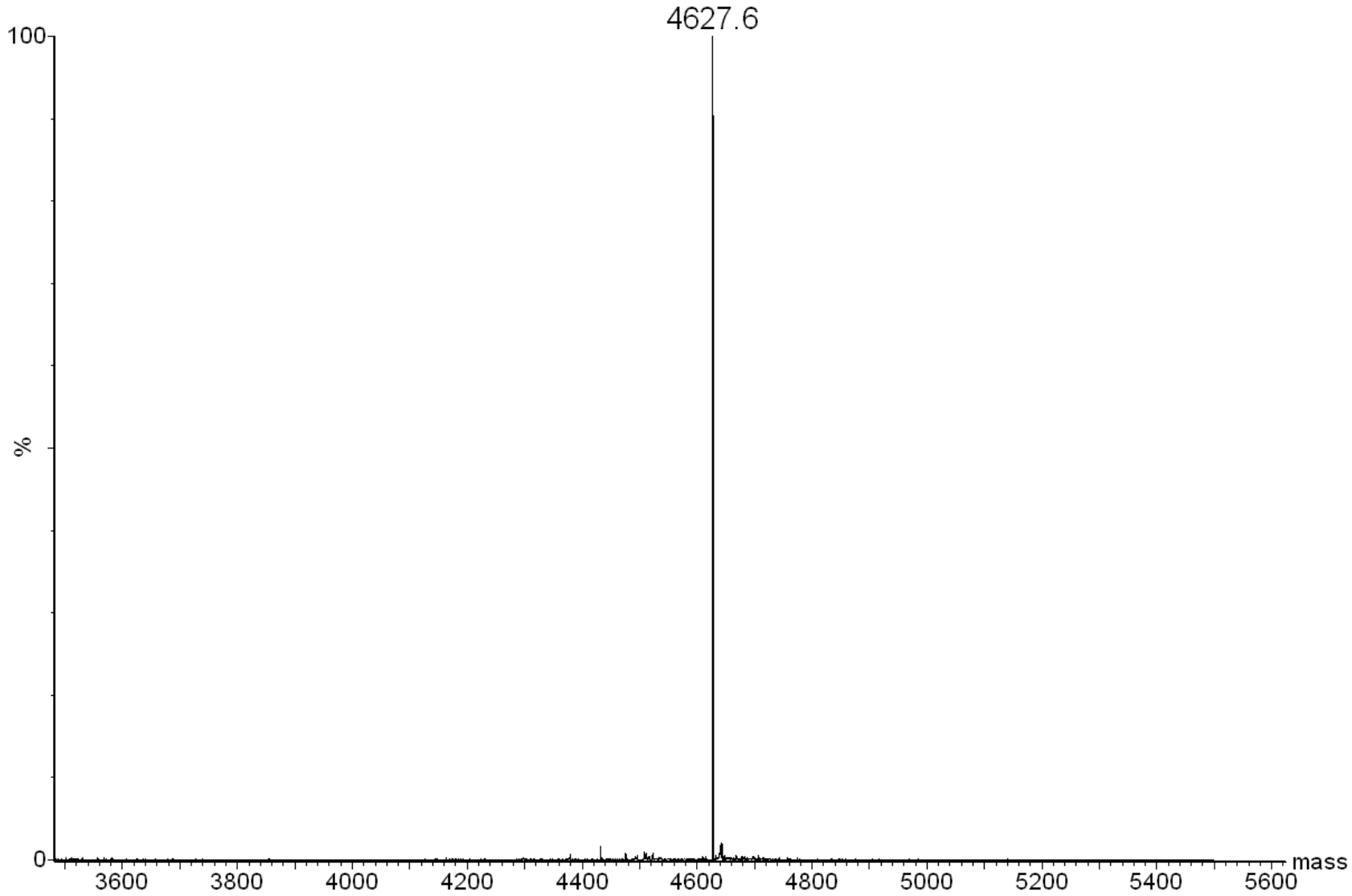
Supplementary Figure 54 - ESI MS spectrum of oligonucleotide **GT7** (expected mass of 4627.2)



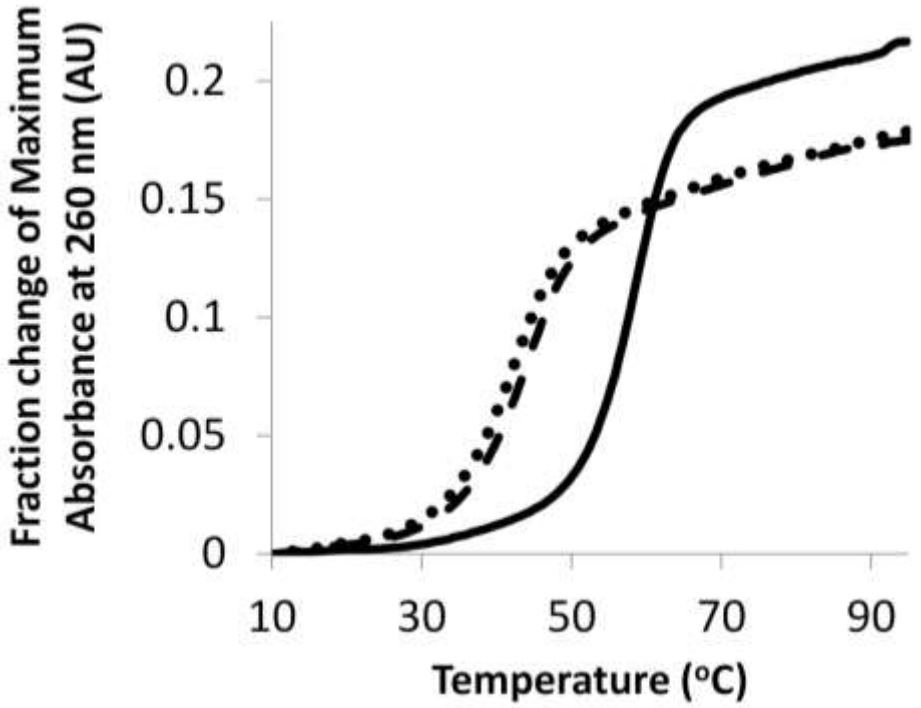
Supplementary Figure 55 - ESI MS spectrum of oligonucleotide **TG4** (expected mass of 4585.1)



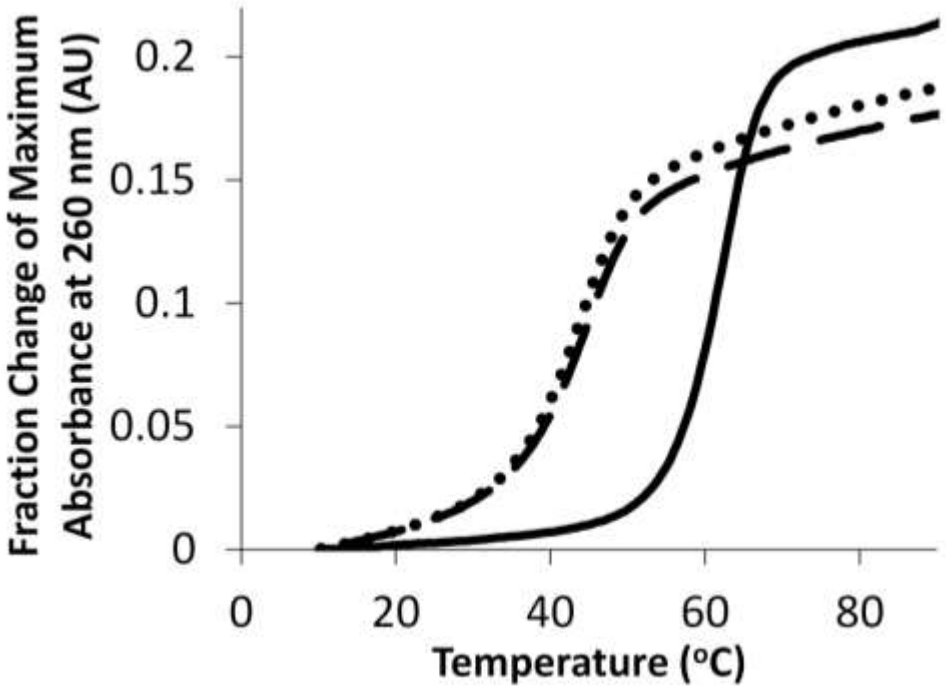
Supplementary Figure 56 - ESI MS spectrum of oligonucleotide **TG7** (expected mass of 4627.2)



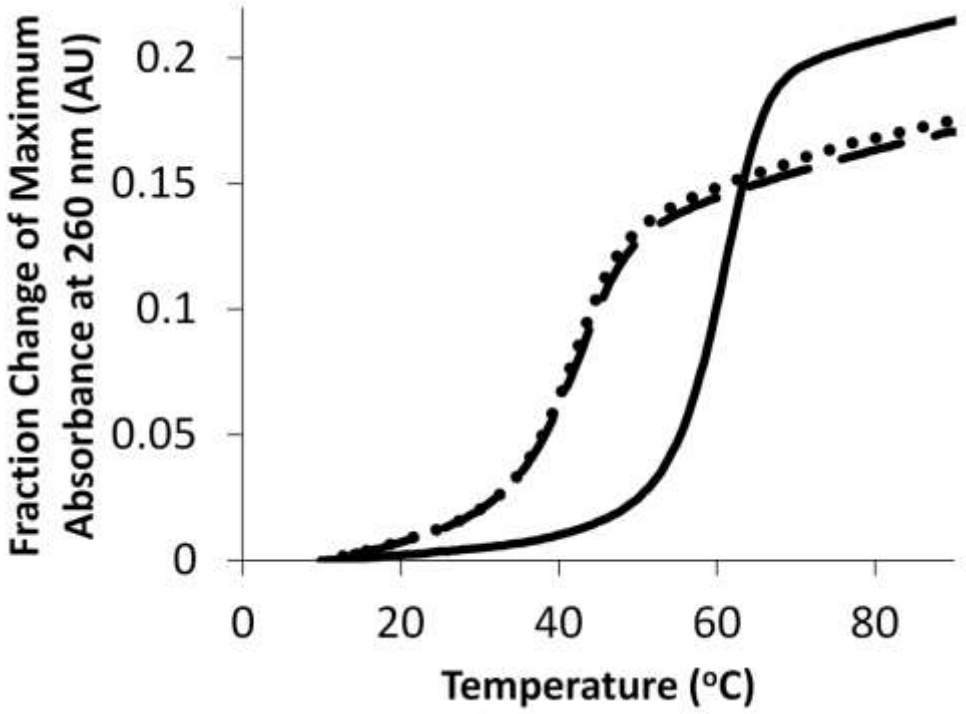
Supplementary Figure 57 - Hyperchromicity change (A_{260}) *versus* temperature (^{o}C) profiles of duplexes containing **TT4** (•••), **TT7** (———) and unmodified **TT** control DNA (—).



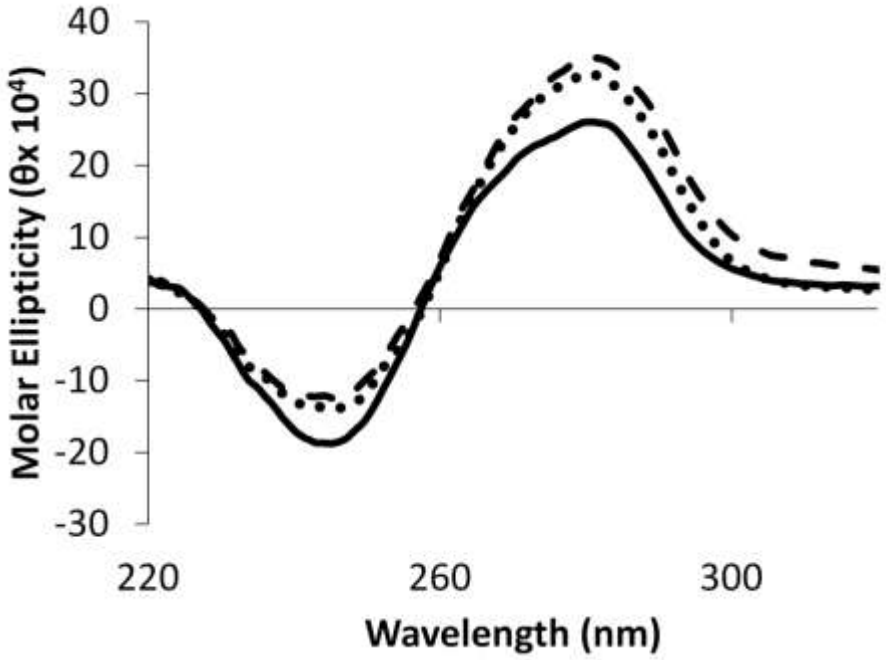
Supplementary Figure 58 - Hyperchromicity change (A_{260}) *versus* temperature (^{o}C) profiles of duplexes containing **GT4** (•••), **GT7** (———) and unmodified **GT** control DNA (—).



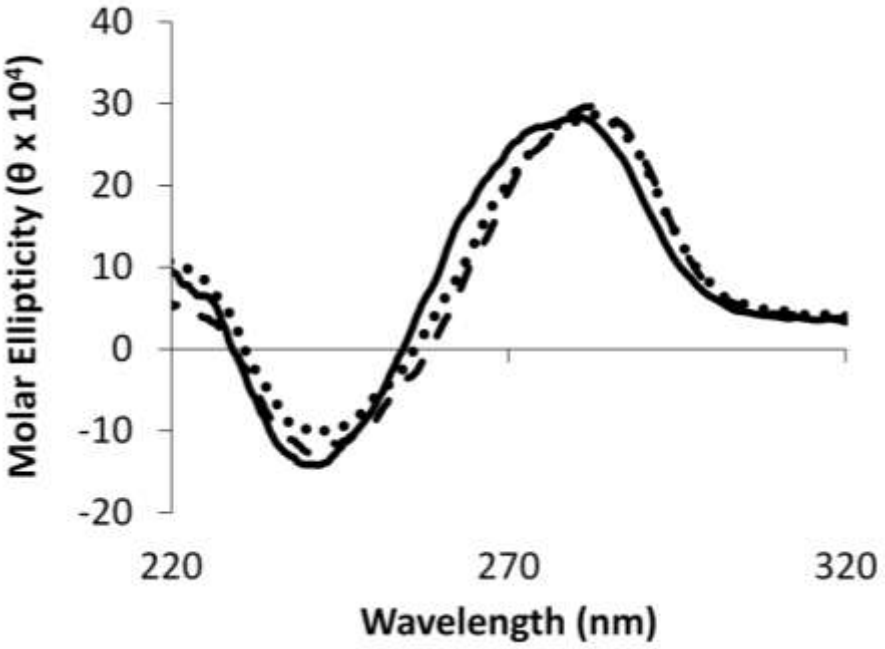
Supplementary Figure 59 - Hyperchromicity change (A_{260}) *versus* temperature (^{o}C) profiles of duplexes containing **TG4** (•••), **TG7** (———) and unmodified **TG** control DNA (—).



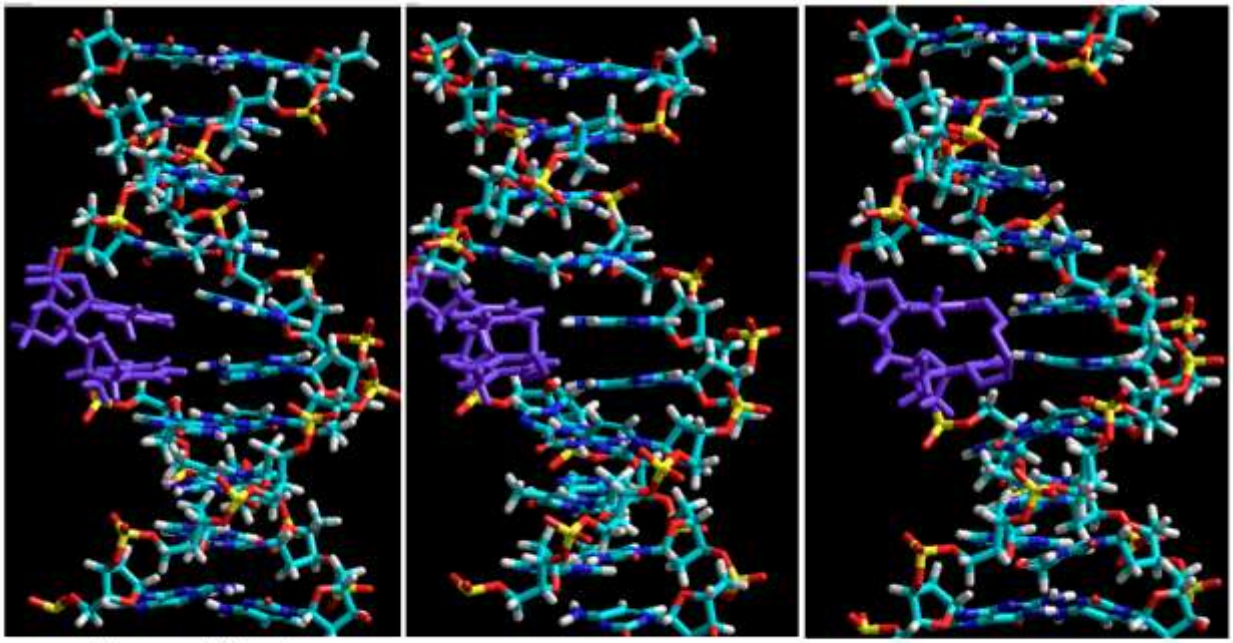
Supplementary Figure 60 - Circular dichroism spectra of IaCL duplexes **TT4** (•••), **TT7** (———) and unmodified control **TT** DNA (—).



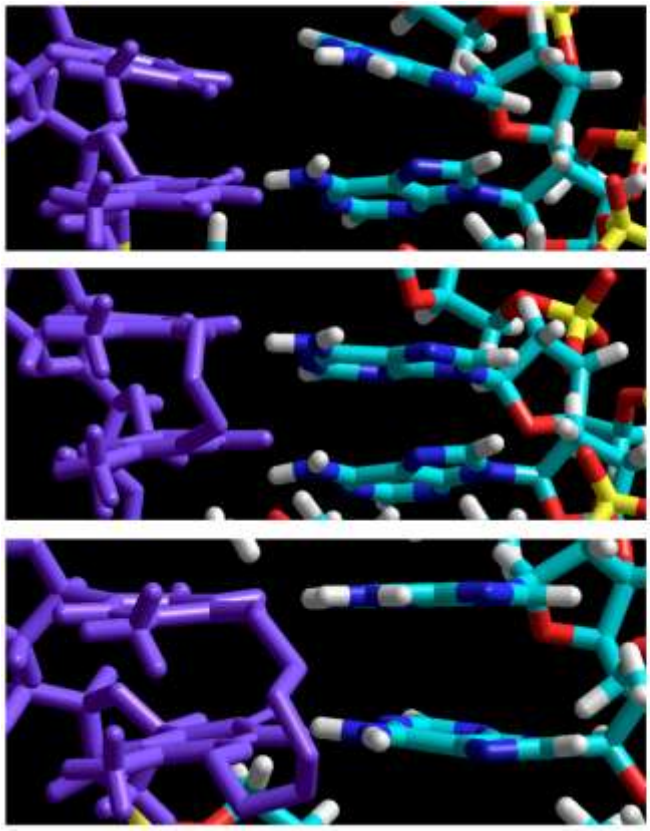
Supplementary Figure 61 - Circular dichroism spectra of IaCL duplexes **GT4** (•••), **GT7** (———) and unmodified control **GT** DNA (—).



Supplementary Figure 62 - Molecular models of unmodified control (**TT**) duplex and duplexes containing **TT4** and **TT7** that were geometry optimized using the AMBER forcefield.



Unmodified control (**TT**) **TT4** **TT7**

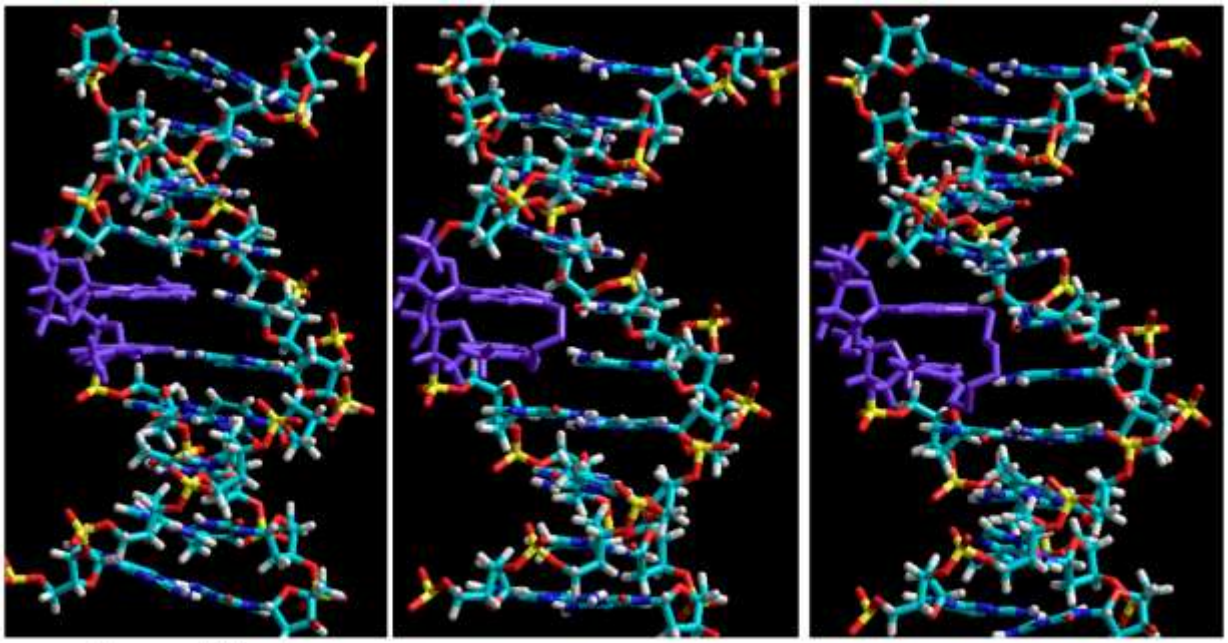


Unmodified control (**TT**)

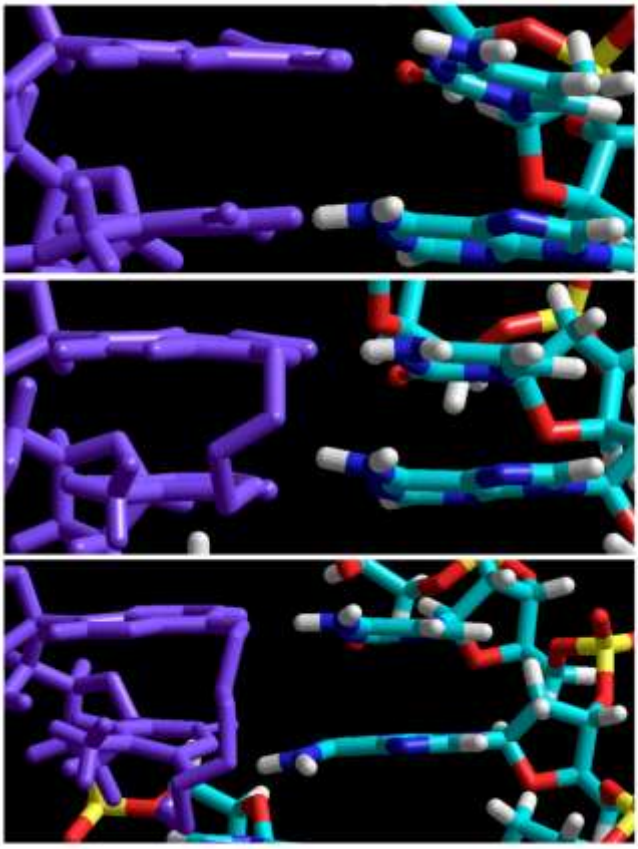
TT4

TT7

Supplementary Figure 63 - Molecular models of unmodified control (**GT**) duplex and duplexes containing **GT4** and **GT7** that were geometry optimized using the AMBER forcefield.



Unmodified control (**GT**) **GT4** **GT7**

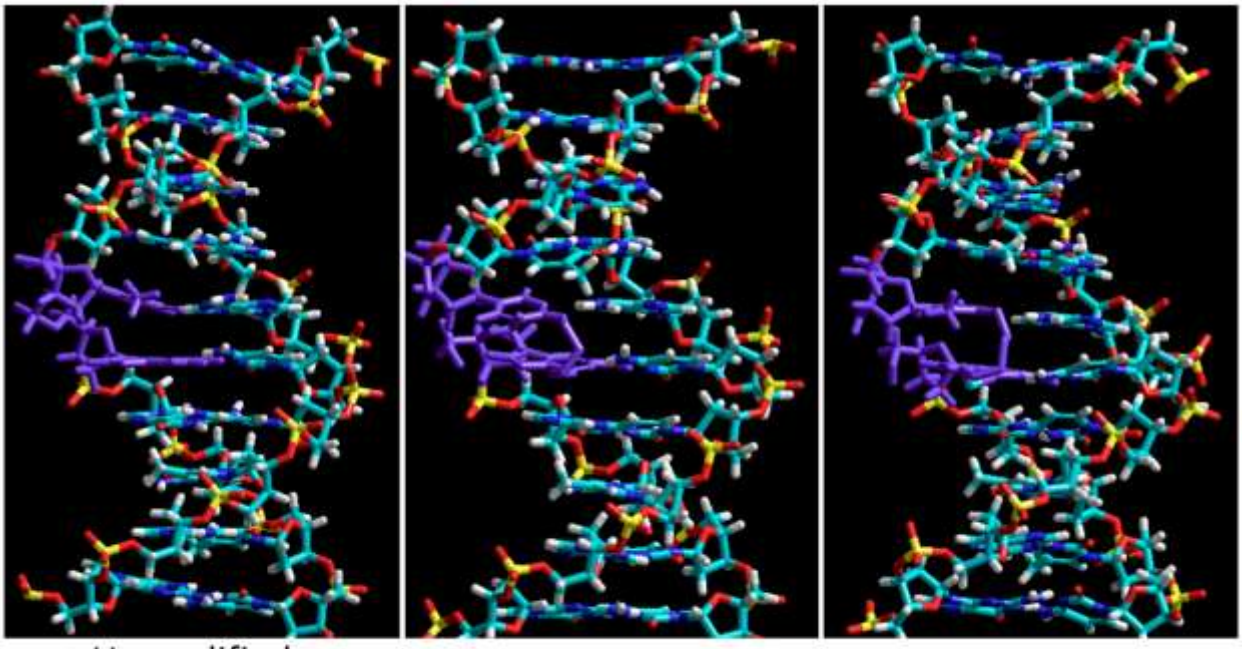


Unmodified control (**GT**)

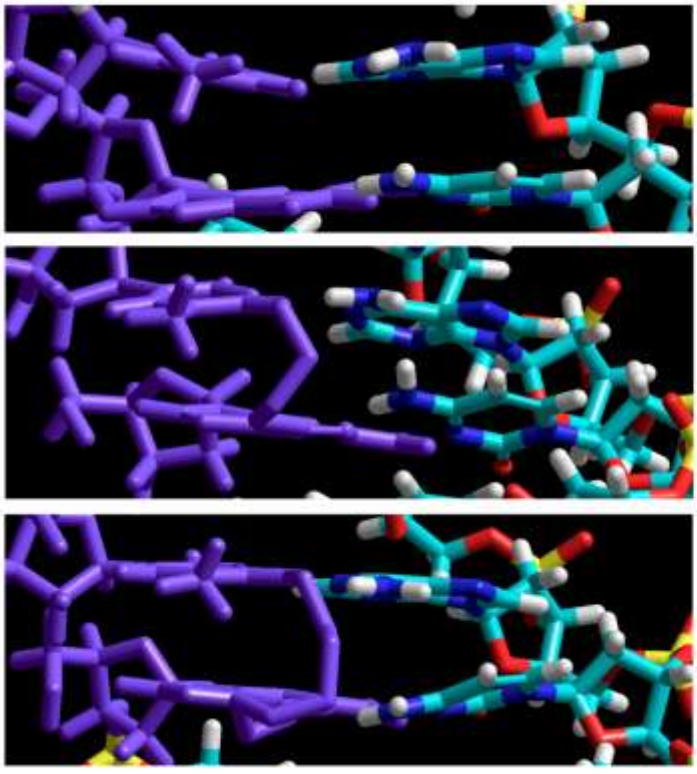
GT4

GT7

Supplementary Figure 64 - Molecular models of unmodified control (**TG**) duplex and duplexes containing **TG4** and **TG7** that were geometry optimized using the AMBER forcefield.



Unmodified control (**TG**) | **TG4** | **TG7**

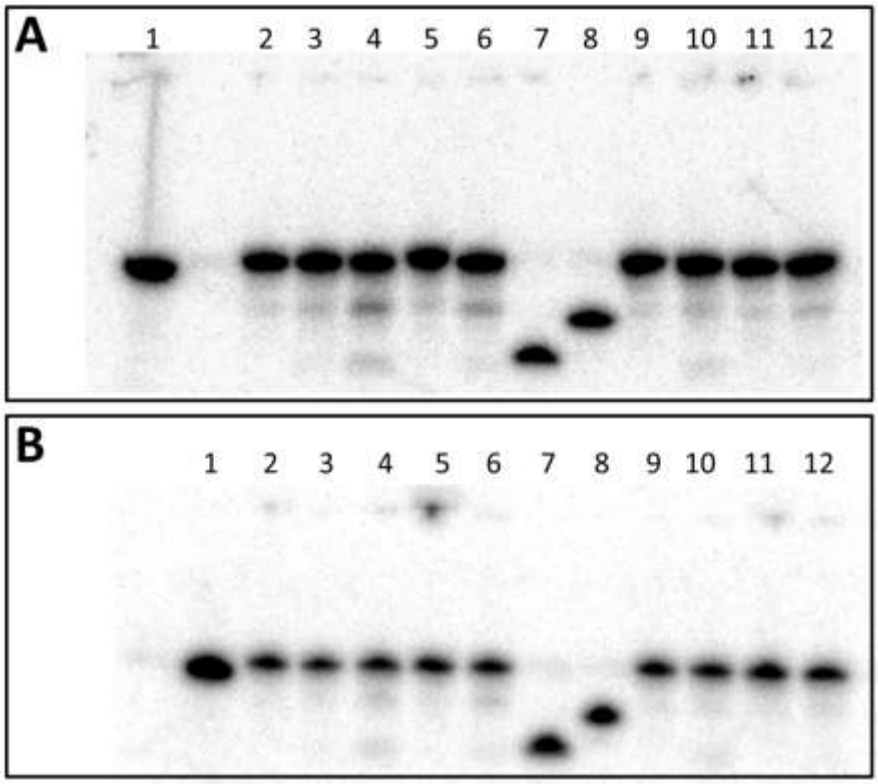


Unmodified control (**TG**)

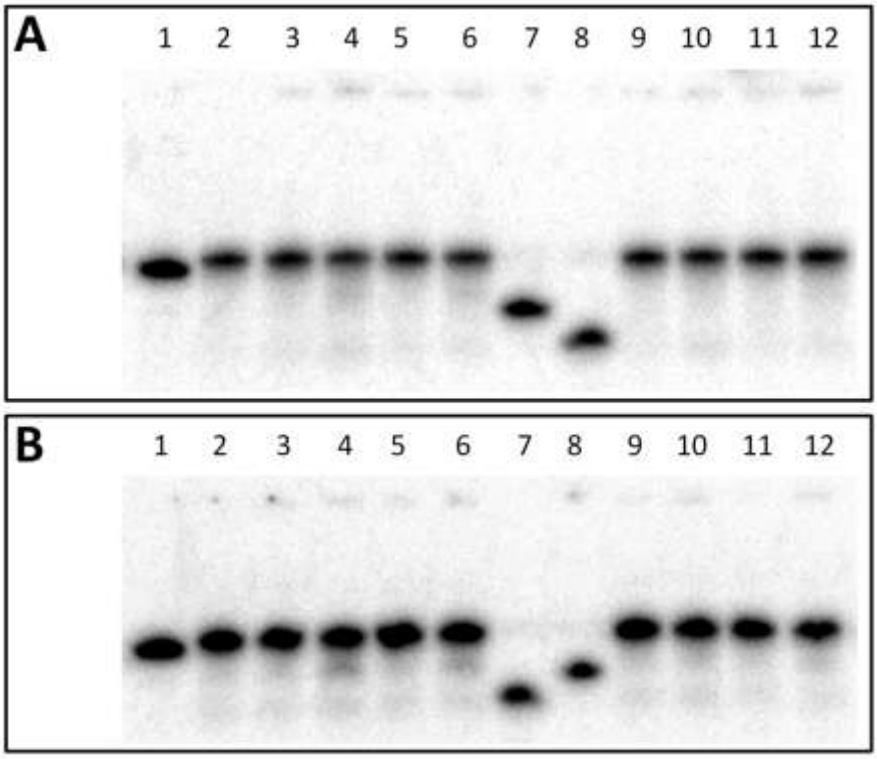
TG4

TG7

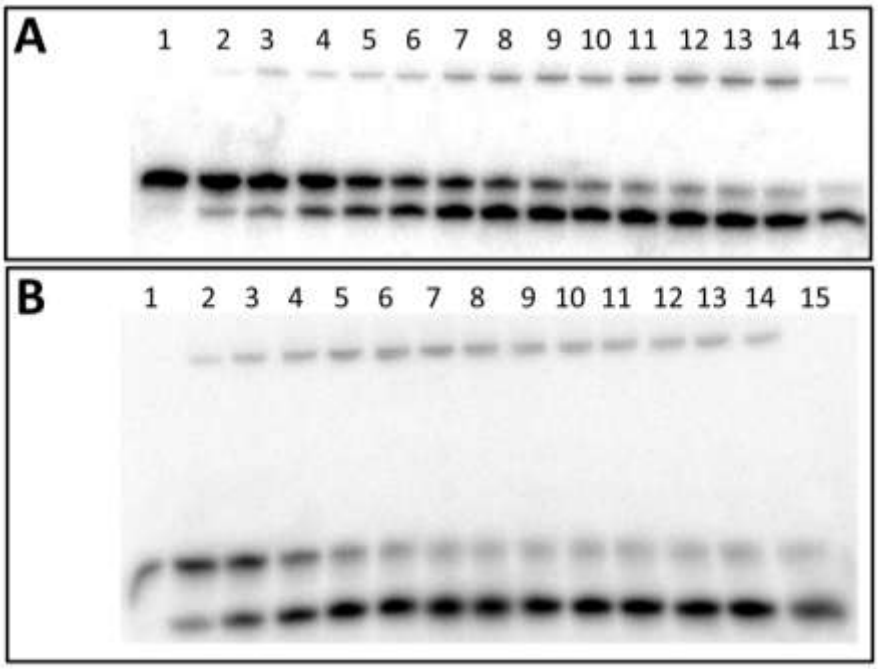
Supplementary Figure 65 - Repair of (**A**) **TT4** and (**B**) **TT7** by hAGT, OGT, Ada-C, and OGT S134P for 16h at 37^{o}C. Denaturing PAGE of repair reactions as described in the experimental section with varying protein concentration (either 5- or 30 fold protein). Panel **A**; Lane 1, 2 pmol control unmodified **TT** DNA; lane 2, 2 pmol **TT4**; lane 3, **TT4** + 10 pmol hAGT; lane 4 **TT4** + 60 pmol hAGT; lane 5 **TT4** + 10 pmol OGT; lane 6 **TT4** + 60 pmol OGT; lane 7 unmodified 5-mer control DNA 5'-pTGCTT; lane 8 unmodified 10-mer control 5'-pTGC ATC ACC AG; lane 9 **TT4** + 10 pmol Ada-C; lane 10 **TT4** + 60 pmol Ada-C; lane 11 **TT4** + 10 pmol OGT S134P; lane 12 **TT4** + 60 pmol OGT S134P. Panel **B**; identical to panel **A** except with **TT7** instead of **TT4**.



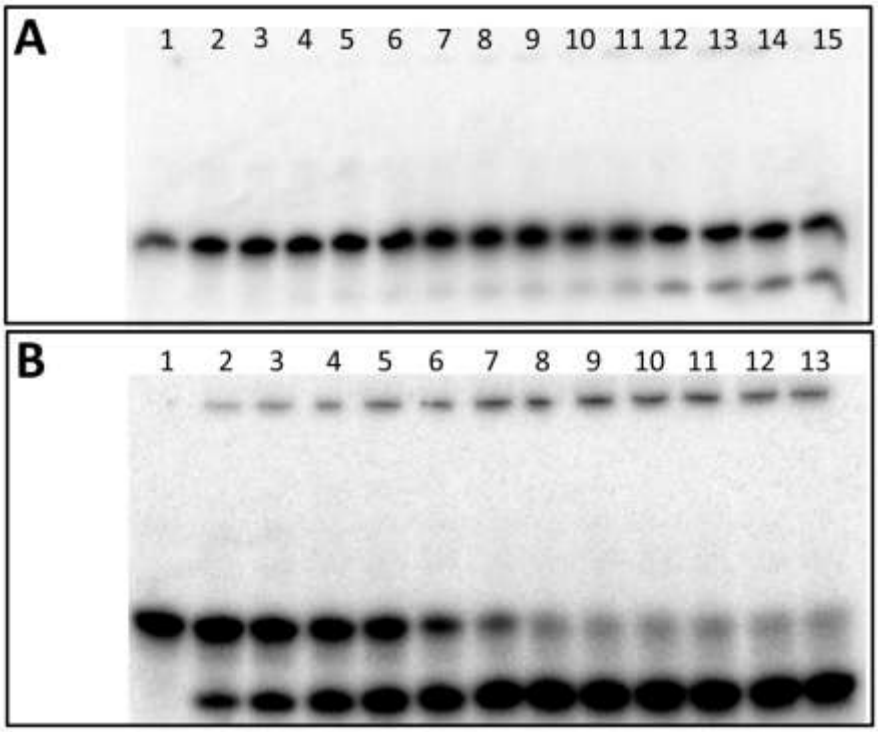
Supplementary Figure 66 - Repair of (**A**) **GT4** and (**B**) **GT7** by hAGT, OGT, Ada-C, and OGT S134P for 16h at 37^{o}C. Denaturing PAGE of repair reactions as described in the experimental section with varying protein concentration (either 5- or 30 fold protein). Panel **A**; Lane 1, 2 pmol control unmodified **GT** DNA; lane 2, 2 pmol **GT4**; lane 3, **GT4** + 10 pmol hAGT; lane 4 **GT4** + 60 pmol hAGT; lane 5 **GT4** + 10 pmol OGT; lane 6 **GT4** + 60 pmol OGT; lane 7 unmodified 5-mer control DNA 5'-pTGCTT; lane 8 unmodified 10-mer control 5'-pTGC ATC ACC AG; lane 9 **GT4** + 10 pmol Ada-C; lane 10 **GT4** + 60 pmol Ada-C; lane 11 **GT4** + 10 pmol OGT S134P; lane 12 **GT4** + 60 pmol OGT S134P. Panel **B**; identical to panel **A** except with **GT7** instead of **GT4**.



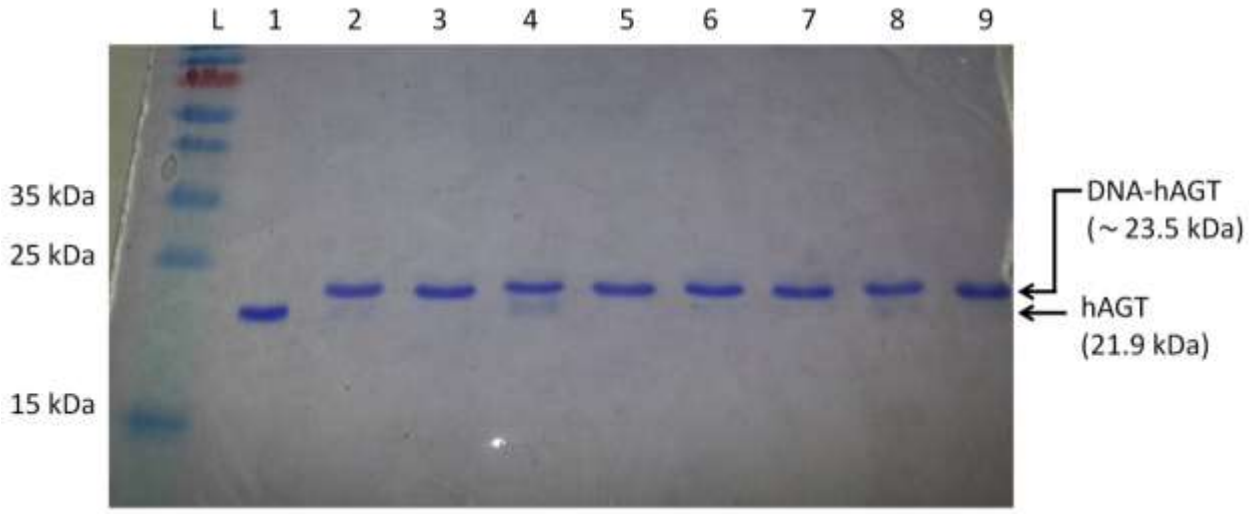
Supplementary Figure 67 - Time course repair gel of duplexes containing (**A**) **TG4** or (**B**) **TG7** by hAGT. (**A**) Denaturing gel of the repair of 2 pmol of **TG4** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **TG4** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240, 1020 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **TG7** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **TG7** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240 and 1020 min, respectively.



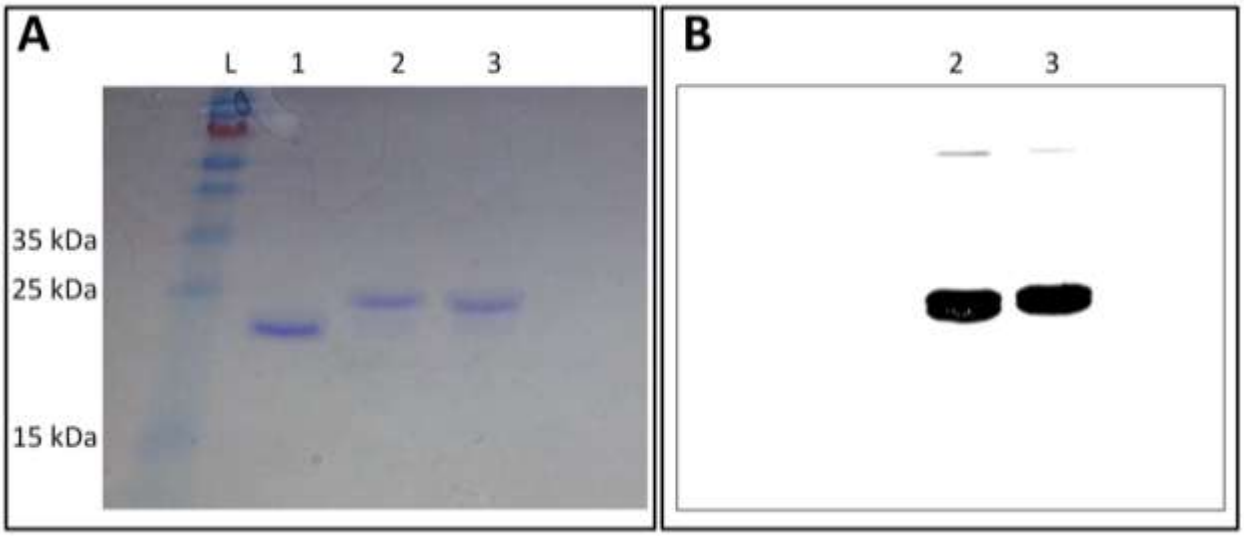
Supplementary Figure 68 - Time course repair gel of duplexes containing **TG7** by 30-fold protein (**A**) Ada-C and (**B**) hAGT. (**A**) Denaturing gel of the repair of 2 pmol of **TG4** by 60 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **TG4** + 60 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240, 1020 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **TG7** by 60 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-13, 2 pmol **TG7** + 60 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2.5, 5, 10, 15, 25, 60, 120 min, respectively.



Supplementary Figure 69 - 15% SDS-PAGE (Coomassie stained) of repair reaction between hAGT (300 pmol) and IaCL DNA (300 pmol) for either 1h (lanes 2-5) or 3.5 h (lanes 6-9) at 37°C. Lane L, molecular weight marker (ladder); lane 1, hAGT control; lane 2, hAGT + **GG4**; lane 3, hAGT + **GG7**; lane 4, hAGT + **TG4**; lane 5, hAGT + **TG7**; lane 6, hAGT + **GG4**; lane 7, hAGT + **GG7**; lane 8, hAGT + **TG4**; lane 9, hAGT + **TG7**. Results demonstrate virtually quantitative conversion to the DPC after only 1h of reaction time.



Supplementary Figure 70 - 15% SDS-PAGE of repair reaction between hAGT (300 pmol) and IaCL DNA (300 pmol) for 1h at 37°C. (**A**) Coomassie Blue stained. (**B**) Fluorescence monitoring of the fluorescein tag. Lane L, molecular weight marker (ladder); lane 1, hAGT control; lane 2, hAGT + **GG4**; lane 3, hAGT + **GG7**. The feasibility of the hAGT protein tagging with a functionalized DNA fragment.



Supplementary References

1 Y. Z. Xu and P. F. Swann, *Nucleic Acids Res.*, 1990, **18**, 4061–4065.

2 F. P. McManus, D. K. O'Flaherty, A. M. Noronha and C. J. Wilds, *Org. Biomol. Chem.*, 2012, **10**, 7078–7090.

3 G. Sun, A. Noronha and C. Wilds, *Tetrahedron*, 2012, **68**, 7787–7793.

4 F. P. McManus, Q. Fang, J. D. M. Booth, A. M. Noronha, A. E. Pegg and C. J. Wilds, *Org. Biomol. Chem.*, 2010, **8**, 4414–4426.

5 F. P. McManus, A. Khaira, A. M. Noronha and C. J. Wilds, *Bioconjugate Chem.*, 2013, **24**, 224–233.

6 D. K. O'Flaherty and C. J. Wilds, *Chem. Eur. J.*, 2015, **21**, 10522–10529.

7 C. J. Wilds, J. D. Booth and A. M. Noronha, *Tetrahedron Lett.*, 2006, **47**, 9125–9128.

8 C. J. Wilds, J. D. M. Booth and A. M. Noronha, *Curr. Protoc. Nucleic Acid Chem*, 2011, **44**, 5.9.1–5.9.19.

9 D. M. Noll, a M. Noronha and P. S. Miller, *J. Am. Chem. Soc.*, 2001, **123**, 3405–3411.

10 J. D. Puglisi and I. J. Tinoco, *Methods Enzym.*, 1989, **180**, 304–325.

11 F. P. McManus and C. J. Wilds, *ChemBioChem*, 2014, **15**, 1966–1977.