Supporting Information

On the enzymatic incorporation of an imidazole nucleotide into DNA

Pascal Röthlisberger,§[a] Fabienne Levi-Acobas,§[a] Ivo Sarac,[a] Philippe Marlière,[b] Piet Herdewijn,[c] and Marcel Hollenstein*[a]

[a]*Laboratory for Bioorganic Chemistry of Nucleic Acids, CNRS, UMR3523, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France. E-mail : marcel.hollenstein@pasteur.fr*

[b]*University of Paris Saclay, CNRS, iSSB, UEVE,Genopole, 5 Rue Henri Desbrueres, 91030 Evry, France.*

[c]KU Leuven, Rega Institute for Medical Research, Medicinal Chemistry, Minderbroedersstraat 10, 3000 Leuven, Belgium.

§ These authors contributed equally.

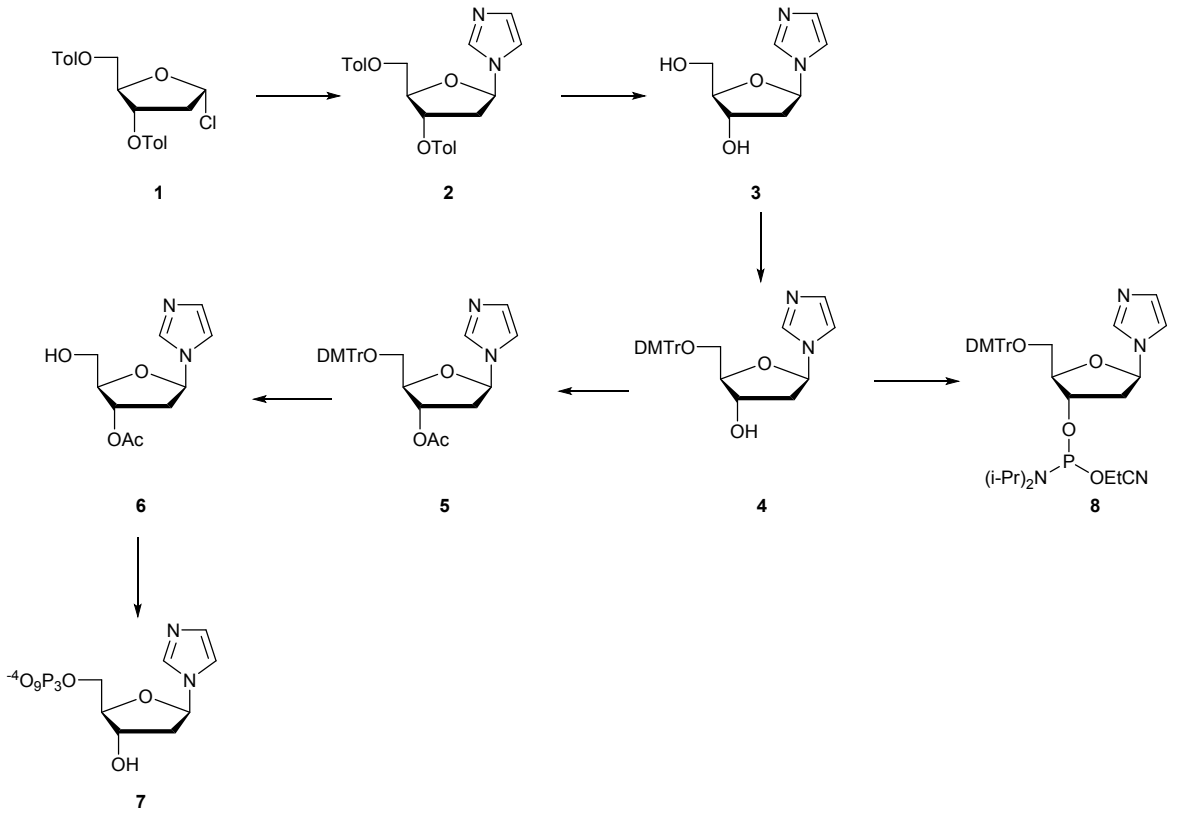
Table of contents

1. General Procedures

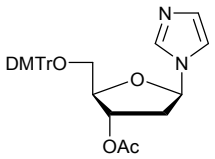
All reactions were performed under N_2 in flame-dried glassware. Anhydrous solvents for reactions were obtained from Sigma Aldrich. Flash chromatography was performed using silica gel (230–400 mesh) from Sigma Aldrich. Thin layer chromatography was carried out on pre-coated glass-backed plates of silica gel (0.25 mm, UV_{254}) from Macherey-Nagel. All chemicals and solvents used were purchased from Sigma-Aldrich and Alfa Aesar. NMR spectra were recorded on a Bruker Avance 400 spectrometer (400.13 MHz for ^{1}H, 100.62 MHz for ^{13}C, and 161.62 MHz for ^{31}P) and all spectra were referenced to the signals of the corresponding solvent. Chemical shifts are given in ppm (δ scale) and coupling constants (*J*) in Hz. Assignation of the NMR signals was performed by using a combination of $^{1}H/^{1}H$-COSY, ^{13}C-DEPT-135, and $^{13}C/^{1}H$-HMBC experiments. High resolution electrospray ionization (ESI) mass spectra (MS, *m*/*z*) were recorded on a Waters Q-Tof Micro MS in the positive-ion electrospray ionization (ESI^{+}) mode. Solutions were prepared using 1:1 MeCN/H_2O containing 0.1% formic acid or MeOH/water containing 10 mM ammonium acetate in the case of sensitive compounds. HPLC purification was performed using an Äkta™ pure system (GE Healthcare) equipped with a Phenomenex Luna semi-preparative RP-HPLC column (5µ C18 100Å). DNA oligonucleotides without imidazole modifications were purchased from Microsynth. DNA oligonucleotides with imidazole modifications were synthesized on an H-8 DNA synthesizer from K&A on a 0.2 µmol scale. Natural DNA phosphoramidites (dT, dC^{4bz}, dG^{2DMF}, dA^{6Bz}) and solid support (dA^{6Bz}-lcaa-CPG 500Å) were all purchased from ChemGenes. Natural DNA phosphoramidites as well as the **dIm** phosphoramidite were prepared as 0.07 M solutions in MeCN and were coupled using 50 sec and 490 sec steps, respectively. 5-(ethylthio)-1*H*-tetrazole (0.25 M in MeCN) was used as coupling agent. Capping, oxidation, and detritylation were performed using standard conditions. Cleavage from the solid support and deprotection of oligonucleotides was achieved by treatment with concentrated ammonia at 55°C for 16 h. After centrifugation, the supernatants were collected and the resulting solutions were evaporated to dryness on a speed-vac. Crude oligonucleotides were purified by anion exchange HPLC (Dionex - DNAPac PA200). Buffer solutions of 25 mM Tris-HCl in H_2O, pH 8.0 (buffer A) and 25 mM Tris-HCl, 1.25 M NaCl in H_2O, pH 8.0 (buffer B) were used. The purified oligonucleotides were then desalted with SepPack C-18 cartridges. Oligonucleotide concentrations were quantitated by UV spectroscopy using a UV5Nano spectrophotometer (Mettler Toledo). The chemical integrity of oligonucleotides was assessed by UPLC-MS analysis: UPLC was performed on a BEH C18 column (130 Å, 1.7 µm, 2.1 mm x 50 mm) from Waters, installed on an ACQUITY UPLC H-Class System (SQ Detector 2). A Buffer containing 20 mM TEA and 400 mM HFIP in H_2O was used with a linear gradient from 18 to 31% Methanol within 5 minutes and a flow rate of 0.3 mL/min.

All the DNA polymerases (Vent (*exo*$^{-}$), *Bst* 2.0, Klenow fragment of DNA polymerase I (*exo*$^{-}$), and Therminator) were purchased from New England Biolabs as well as the natural dNTPs. Acrylamide/bisacrylamide (29:1, 40%) was obtained from Fisher Scientific. Visualization of PAGE gels was performed by fluorescence imaging using a Storm 860 phosphorimager with the ImageQuant software (both from GE Healthcare).

2. Synthesis of nucleoside and nucleotide analogues

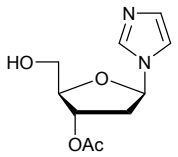


Scheme S1. Schematic representation of the nucleoside synthesis



(150 mg, 0.31 mmol) of the trytilated nucleoside **4** was dissolved in dry pyridine (2 mL) at RT under N_2. To this solution DMAP (12.5 mg, 0.10 mmol, 0.33 eq.), Et_3N (143 µL, 1.03 mmol, 3.3 eq.) and acetic anhydride (46 µL, 0.49 mmol, 1.6 eq.) were added. After stirring for 3h at this temperature, the reaction mixture was quenched by the addition of $NaHCO_3$ sat. (10 mL) and then extracted with DCM (3x 20 mL). The organic layers were combined, dried over $MgSO_4$, concentrated and purified by FC (DCM/MeOH 1-5%) to yield 160 mg (quant.) of **5** as a yellow oil.

^{1}H-NMR ($CDCl_3$) δ = 2.11 (1H, s), 2.48 (1H, m, *J* = 3.02 Hz), 2.65 (1H, m, *J* = 4.79 Hz), 3.38 (1H, d, *J* = 3.92 Hz), 3.81 (1H, d, *J* = 1.04 Hz), 4.21 (1H, m, *J* = 1.95 Hz), 5.43 (1H, m, *J* = 1.92 Hz), 6.03 (1H, q, *J* = 4.71 Hz), 6.84 (1H, m, *J* = 2.10 Hz), 7.07 (1H, s), 7.26 (1H, m, *J*=4.71 Hz), 7.43 (1H, m, *J*=2.42 Hz), 7.67 (1H, s). ^{13}C-NMR ($CDCl_3$) δ = 21.0, 21.4 , 39.2 , 55.2 , 63.7 , 75.1 , 84.0 , 86.1 , 86.7 , 113.2 , 116.5 , 125.3 , 127.0 , 127.9 , 128.2 , 128.2 , 129.0 , 130.1 , 130.1 , 135.6 , 136.2 , 137.9 , 144.5 , 158.6 , 170.2 . ESI MS $C_{31}H_{33}N_2O_6^+$ *m/z* calc. 529.2333, obs. 529.2360.



Tritylated nucleoside **5** (190 mg, 0.36 mmol) was dissolved in dry chloroform (5 mL) at RT under an N_2 atmosphere. To this solution, TFA (276 µL, 3.6 mmol, 10 eq.) was added. The orange solution was stirred for 20 min at rt then quenched with $NaHCO_3$ sat. (10 mL) and extracted with DCM (3 x 15 mL). The combined organic layers were dried over $MgSO_4$, concentrated under reduced pressure and finally purified by FC (DCM/ MeOH 5-10%) to yield 63 mg (78%) of the desired product **6**.

^{1}H-NMR (MeOD) δ = 2.11 (1H, s), 2.58 (1H, m), 3.76 (1H, d, *J* = 3.72 Hz), 4.13 (1H, m), 5.36 (1H, m), 6.13 (1H, dd, *J* = 5.74, 8.34 Hz), 7.01 (1H, s), 7.40 (1H, s), 7.92 (1H, s). ^{13}C-NMR (MeOD) δ = 19.5 , 38.7 , 61.8 , 75.1 , 85.5 , 86.4 , 117.1 , 128.1 , 136.2 , 170.8. ESI MS $C_{10}H_{15}N_2O_4^+$ *m/z* calc. 227.1026, obs. 227.1026. UV/Vis λ_{max} =209 nm, ε =2828 $Lmol^{-1}cm^{-1}$

3. Additional primer extension reactions

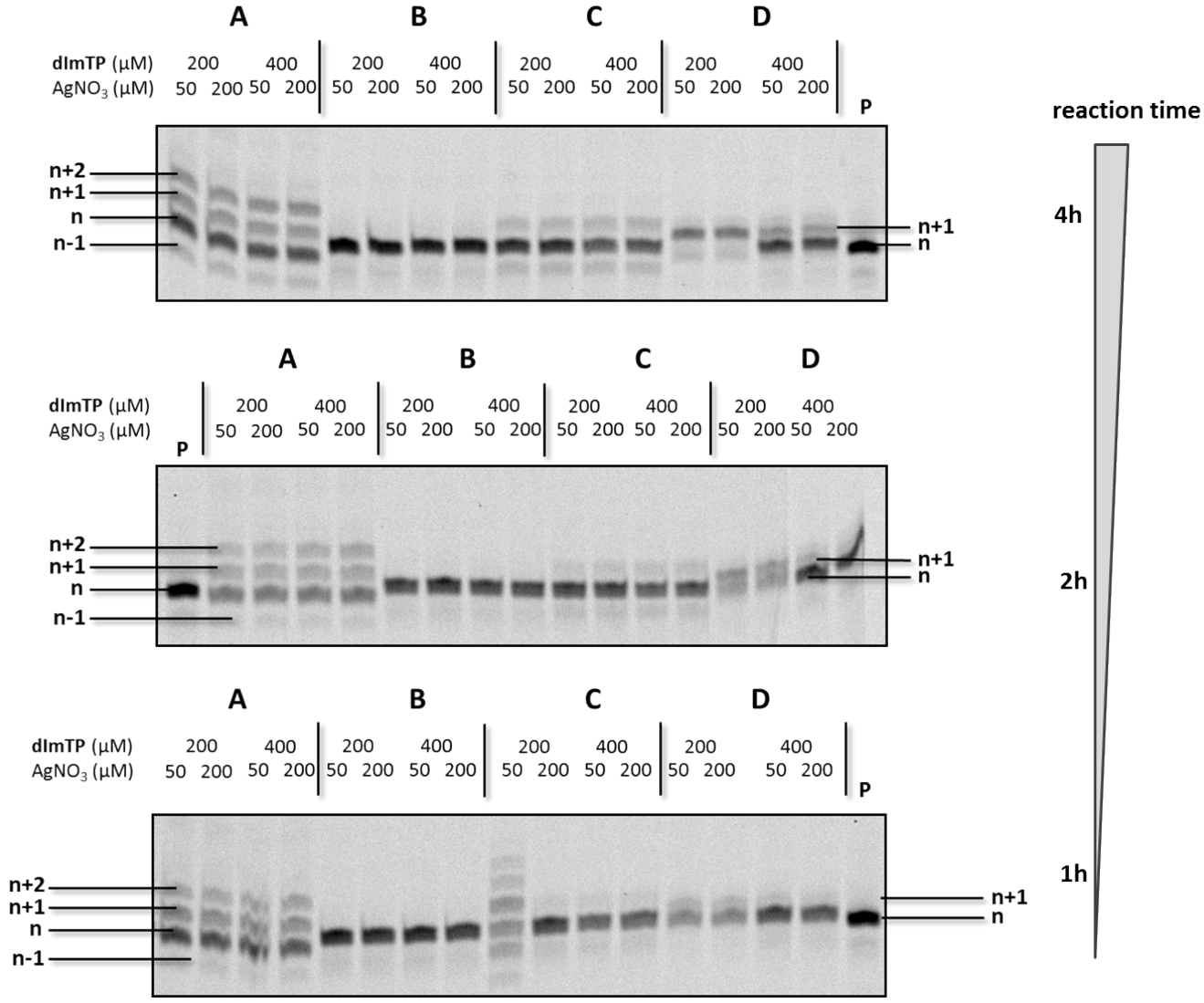


Figure S1. PAGE (20%) analysis of single **dImMP** insertions opposite **dIm** using the Therminator (**A**), Vent (*exo*$^-$) (**B**), *Bst* (**C**), and the Kf *exo*$^-$ (**D**) DNA polymerases with primer **P1** and template **T1**. The reaction mixtures contained the indicated concentrations of **dImTP** and $AgNO_3$, DNA polymerase (2U of Therminator and Vent (*exo*$^-$) and 5U of Kf *exo*$^-$ and 8U of *Bst*) and were incubated for various amounts of time (1h, 2h, and 4h). **P** indicates the unreacted primer.

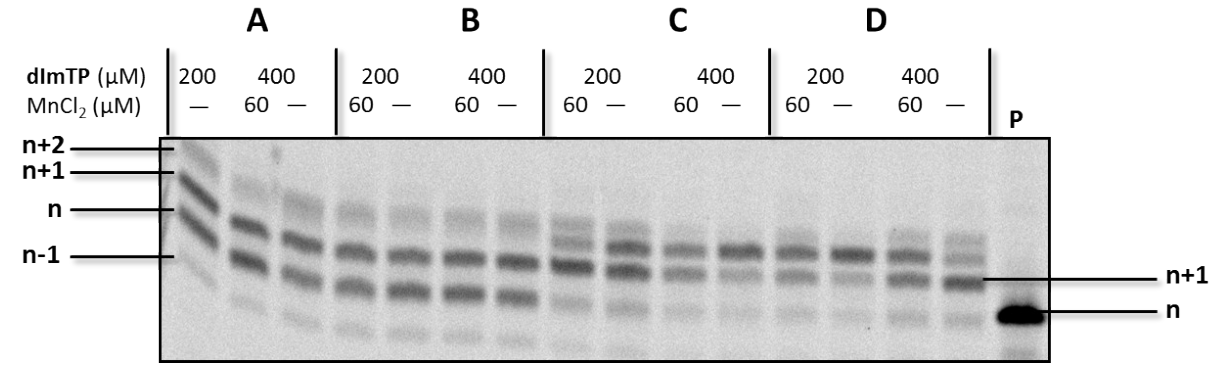


Figure S2. PAGE (20%) analysis of primer extension reactions with primer **P1** and template **T1** (A and B) and primer **P1** and template **T6** (C and D) with the Therminator DNA polymerase. All reactions contained given amounts of **dImTP** and $MnCl_2$, 2U of polymerase and were incubated at 60°C for 1h (A and C) or 2h (B and D). **P** indicates the unreacted primer.

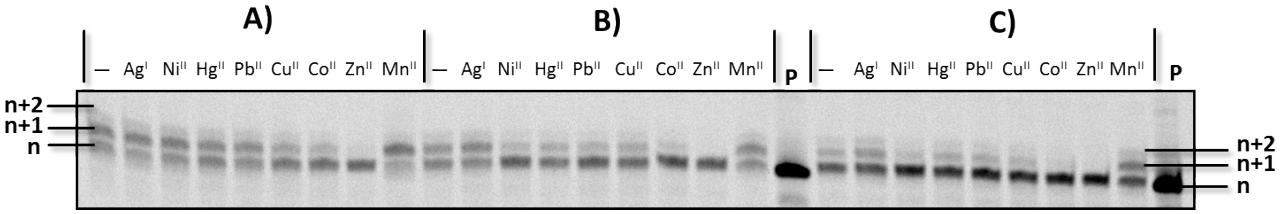


Figure S3. Gel analysis (PAGE 20%) of the effect of the nature of the M^{n+} on the PEX with **dImMP** and templates **T1** A), **T7** B), and **T8** C). The reaction mixtures contained 200 μM **dImTP**, 50 μM metal cations, and 5U of polymerase and were incubated at 37°C for 6h in 1x reaction buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, **10 μM DTT**, pH 7.9). **P** indicates the unreacted primer.

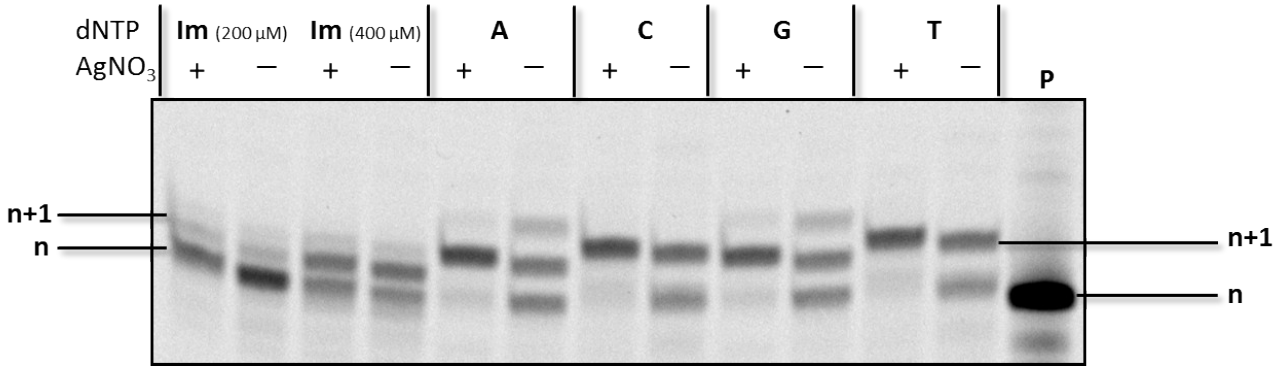


Figure S4. Gel image (PAGE 20%) of PEX with primer **P1** and template **T1** with Kf *exo*$^{-}$ in the presence or absence of $AgNO_3$ and with the different natural dNTPs as well as **dImTP**. The reaction mixtures contained 10 μM natural dNTPs or 200/400 μM **dImTP**, 15 μM $AgNO_3$, and 5U of polymerase and were incubated at 37°C for 6h. **P** indicates the unreacted primer.

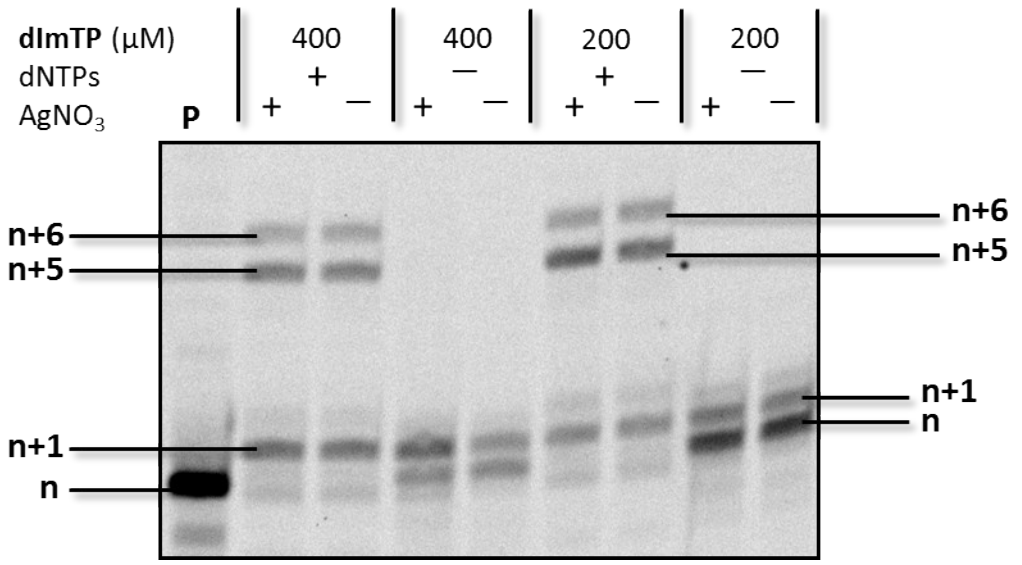


Figure S5. Gel analysis (PAGE 20%) of the experiment designed to yield full replication of template **T1**. A) PEX with primer **P1** and template **T1** with 200/400 μM **dImTP** and in the presence or absence of 100 μM $AgNO_3$. B) The reaction products stemming from A) were incubated in the presence of 5U of Kf *exo*$^-$ and the four canonical dNTPs (20 μM) for 1h at 37°C. **P** indicates the unreacted primer.

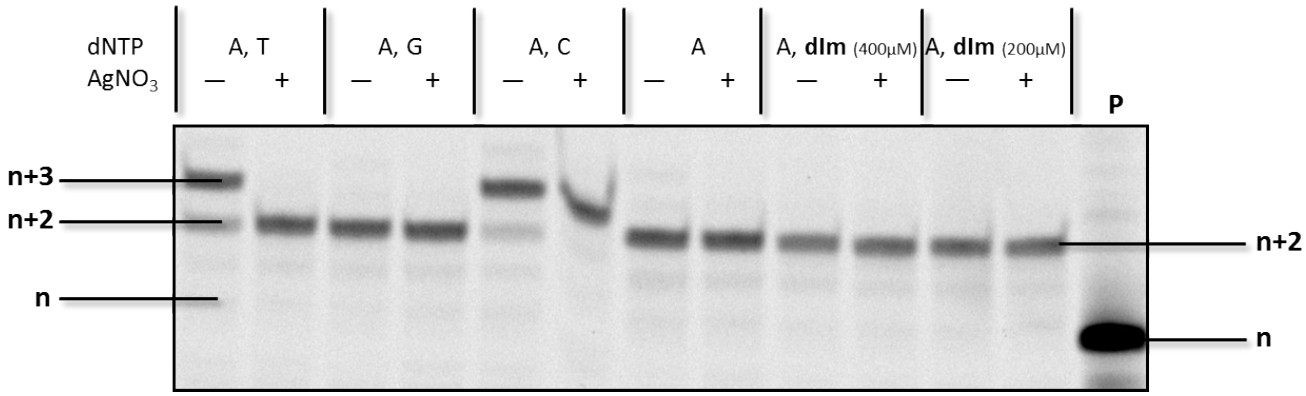


Figure S6. Gel image (PAGE 20%) of PEX with primer **P1** and template **T6** with Kf *exo*$^-$ in the presence or absence of $AgNO_3$ and with the different natural dNTPs as well as **dImTP**. The reaction mixtures contained 10 μM natural dNTPs or 200/400 μM **dImTP**, 15 μM $AgNO_3$, and 5U of polymerase and were incubated at 37°C for 6h. **P** indicates the unreacted primer.

4. Copies of NMR and ESI-MS spectra

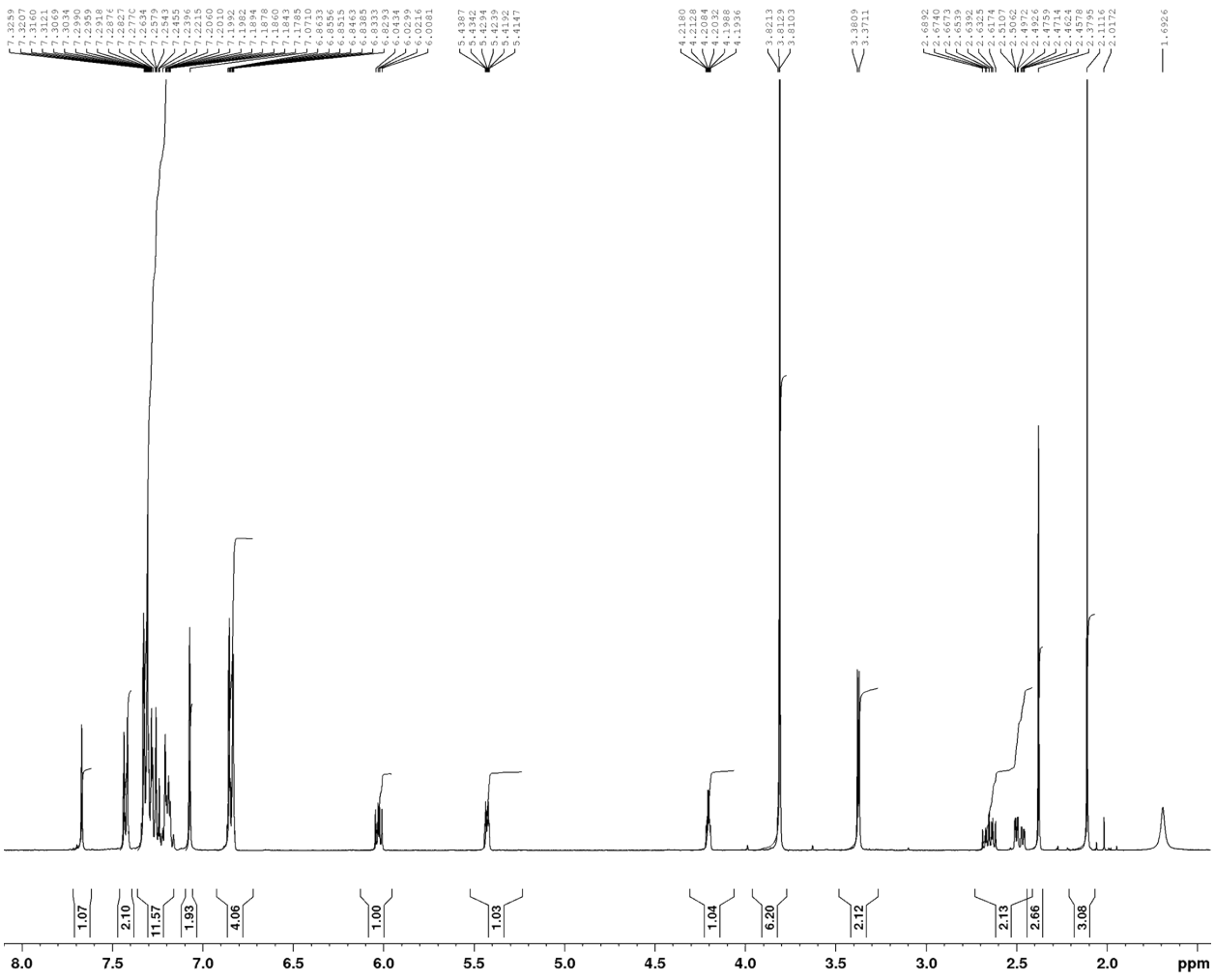


Figure S7. ^{1}H-NMR ($CDCl_3$) spectrum of **5**

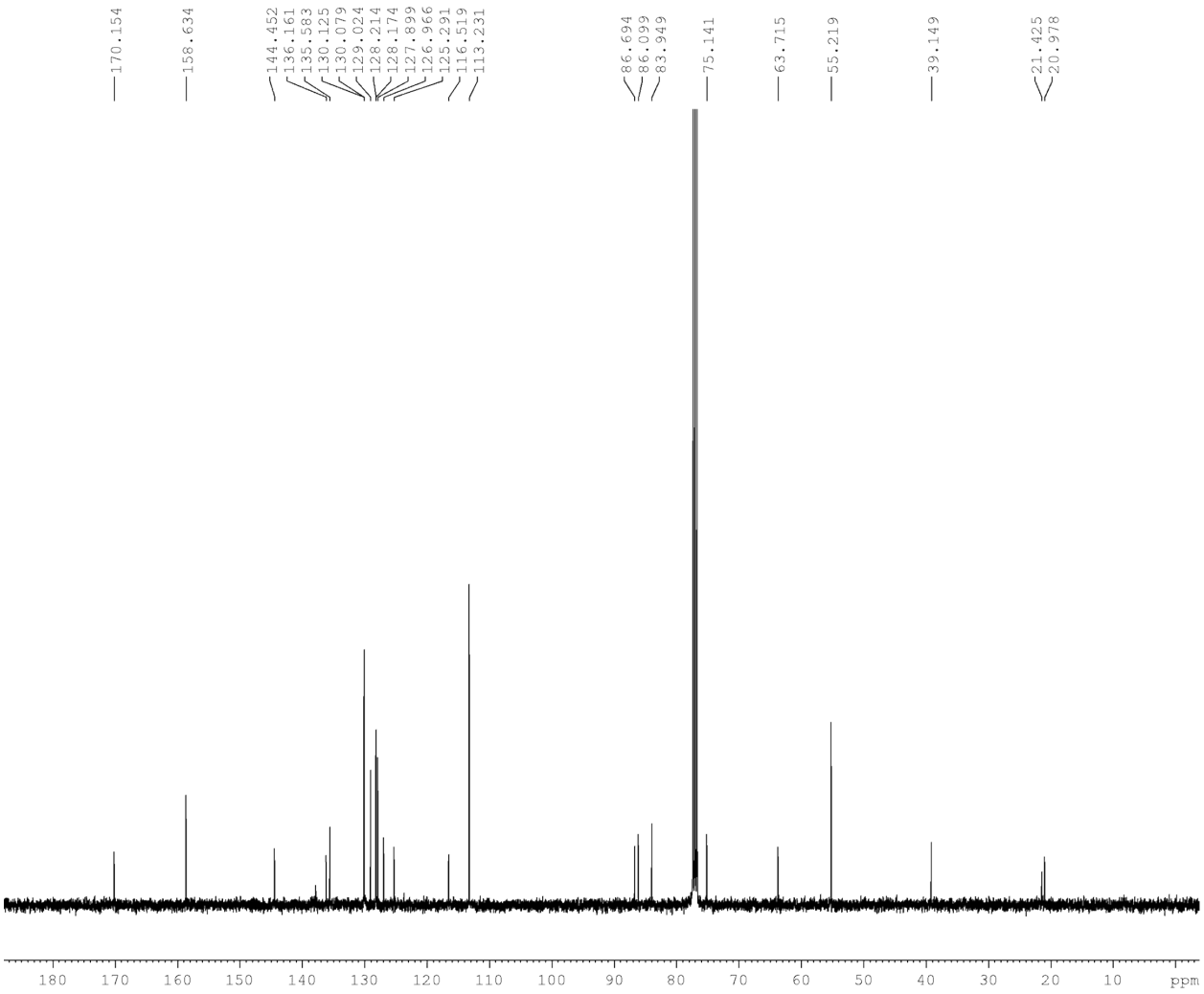


Figure S8. ^{13}C-NMR ($CDCl_3$) spectrum of **5**

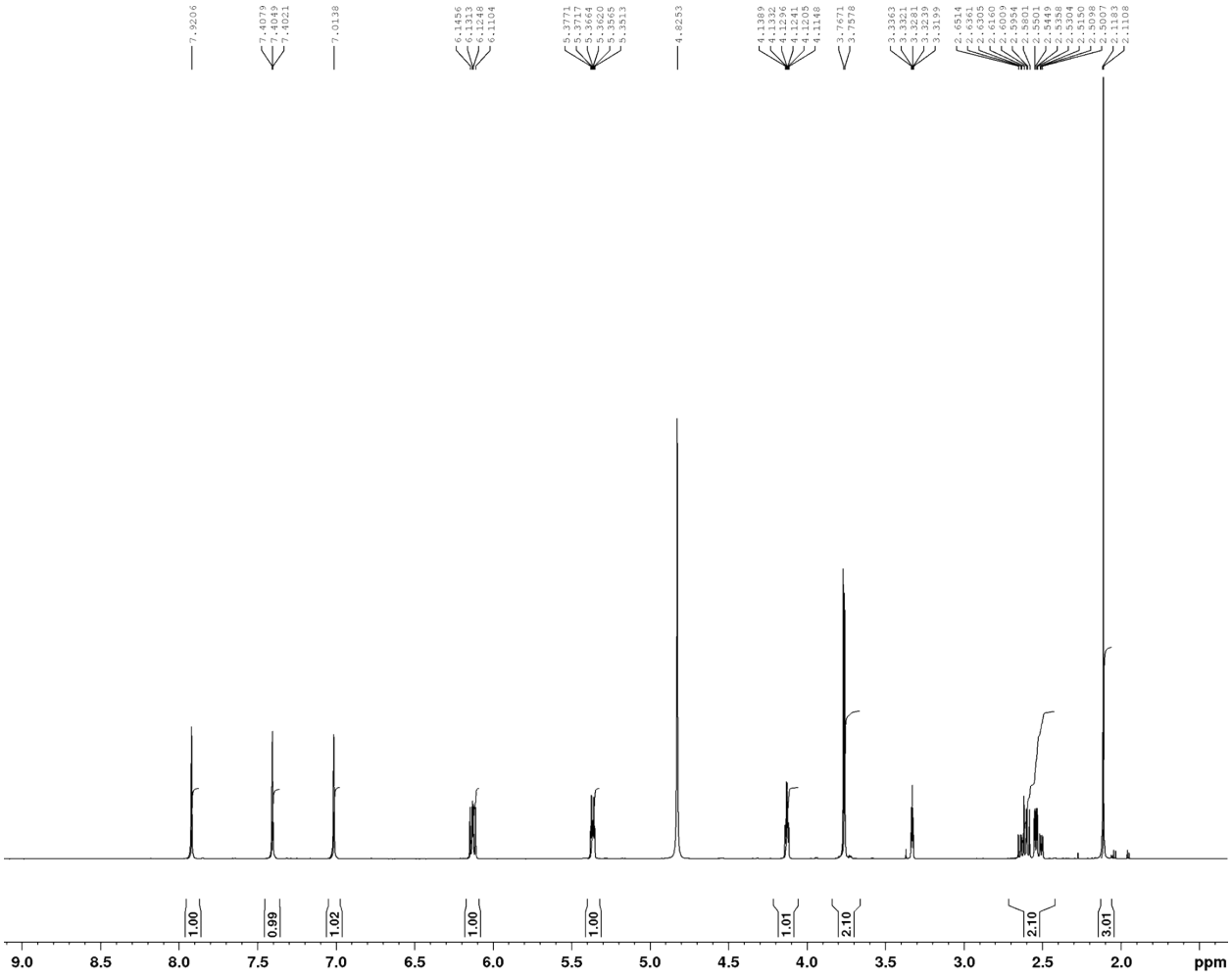


Figure S9. ^{1}H-NMR (MeOD) spectrum of **6**

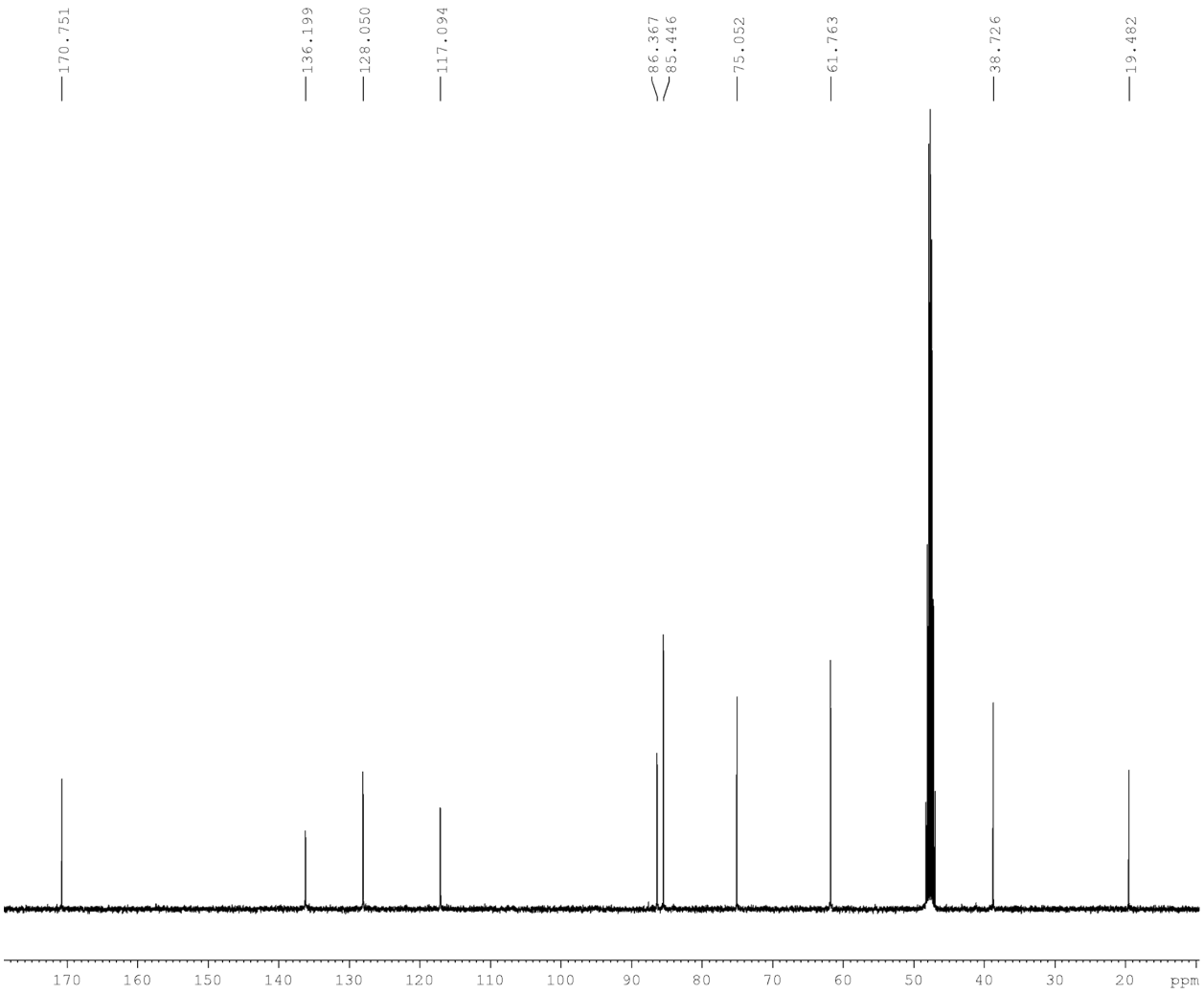


Figure S10. ^{13}C-NMR (MeOD) spectrum of **6**

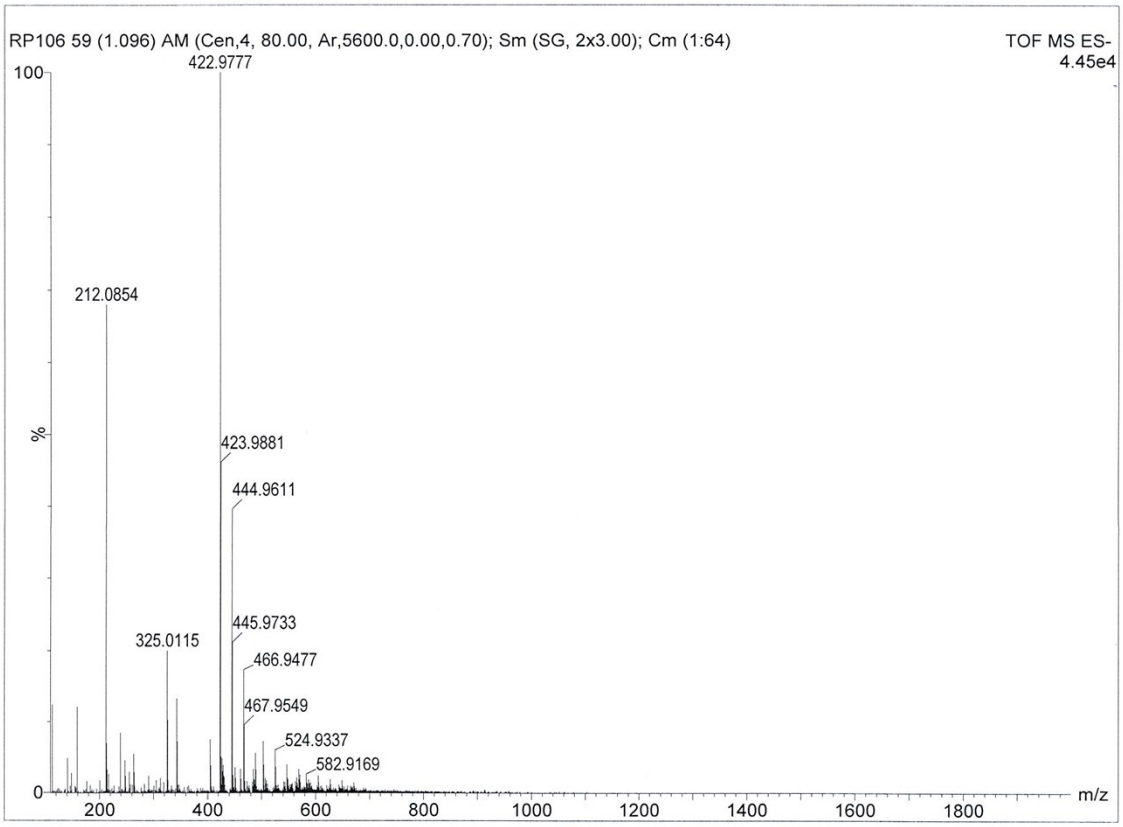


Figure S11. ESI-MS analysis of **7**

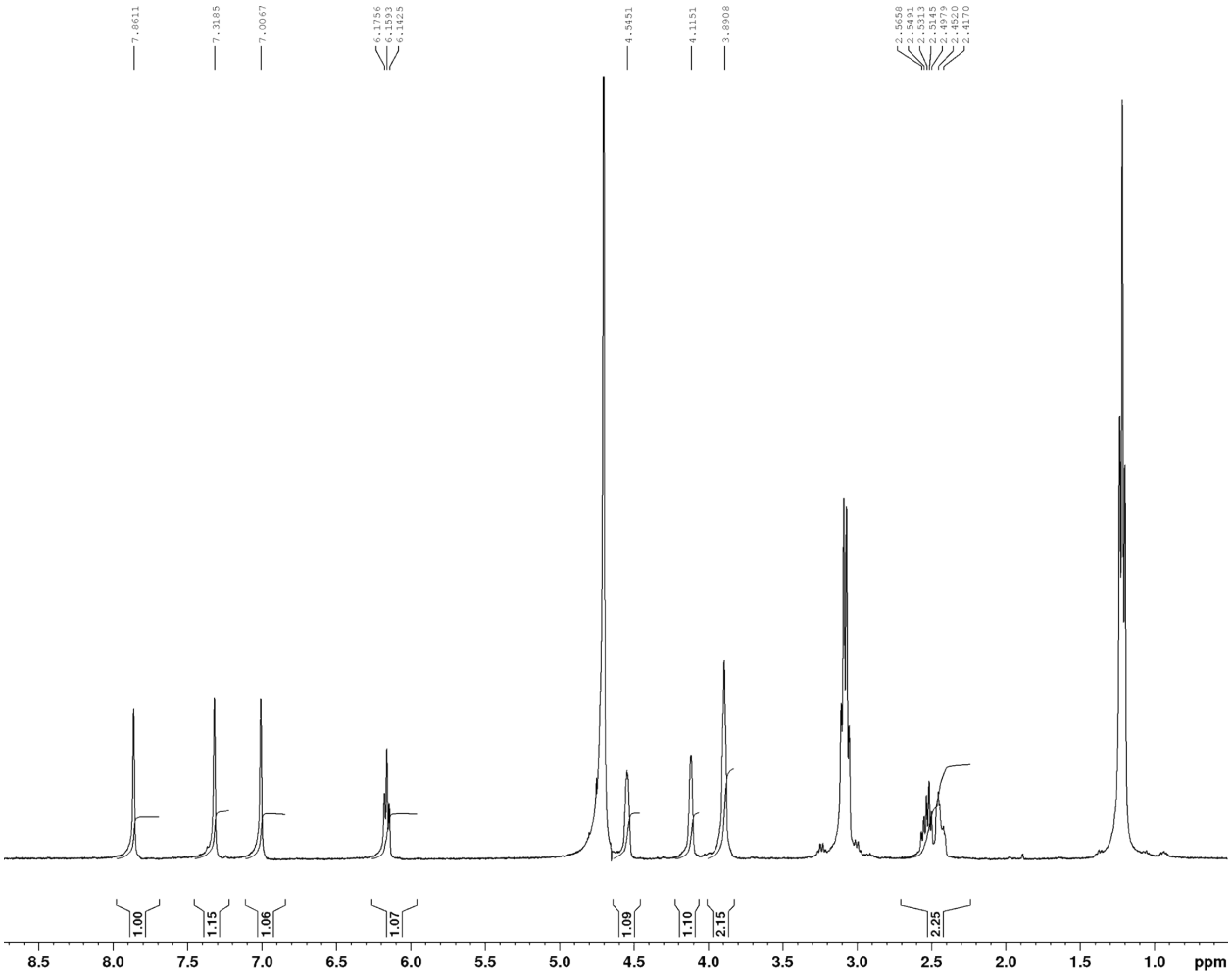


Figure S12. ^{1}H-NMR (D_2O) spectrum of **7**

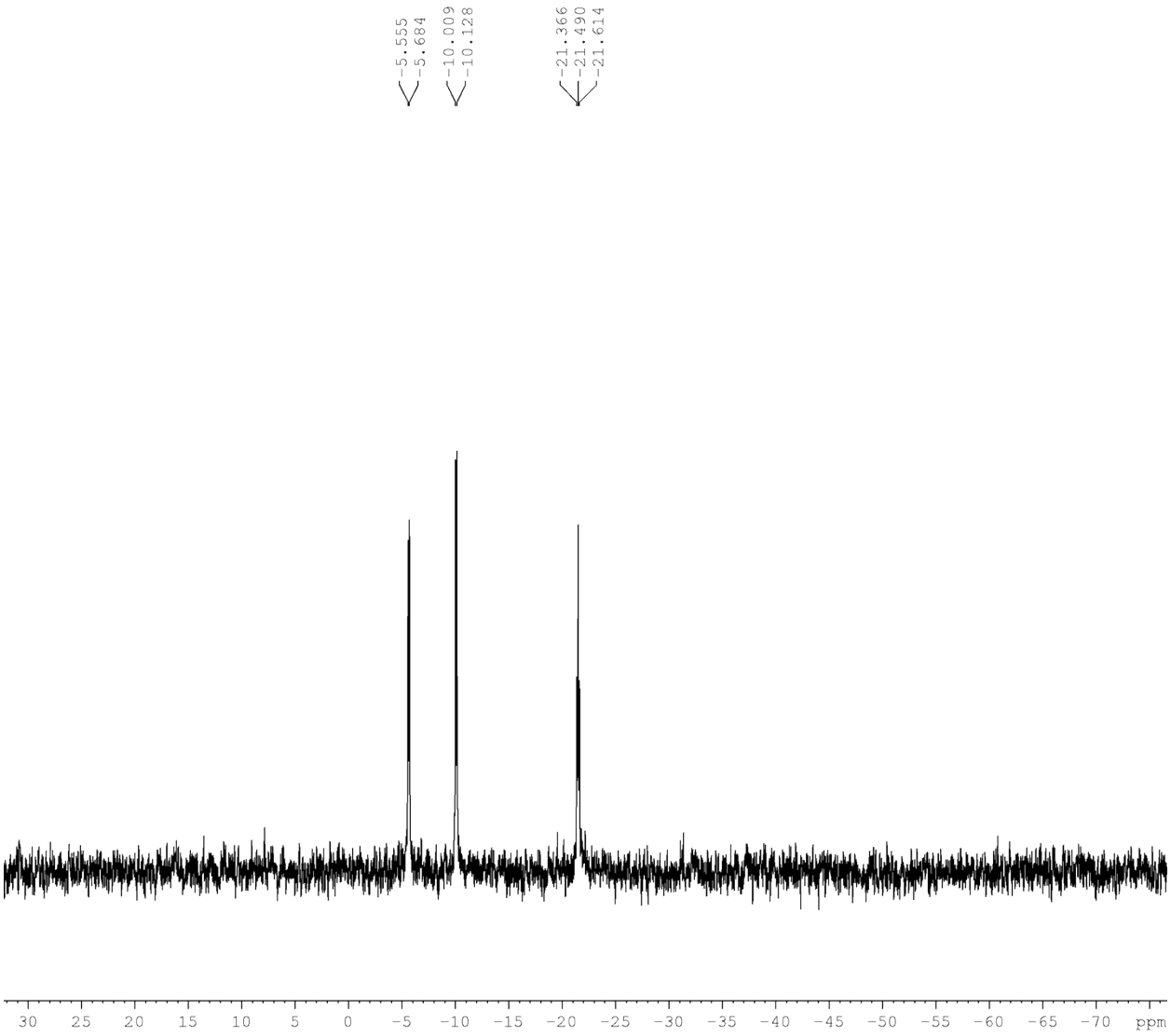


Figure S13. ^{31}P-NMR (D_2O) spectrum of **7**

5. UPLC-MS analysis of modified oligonucleotides

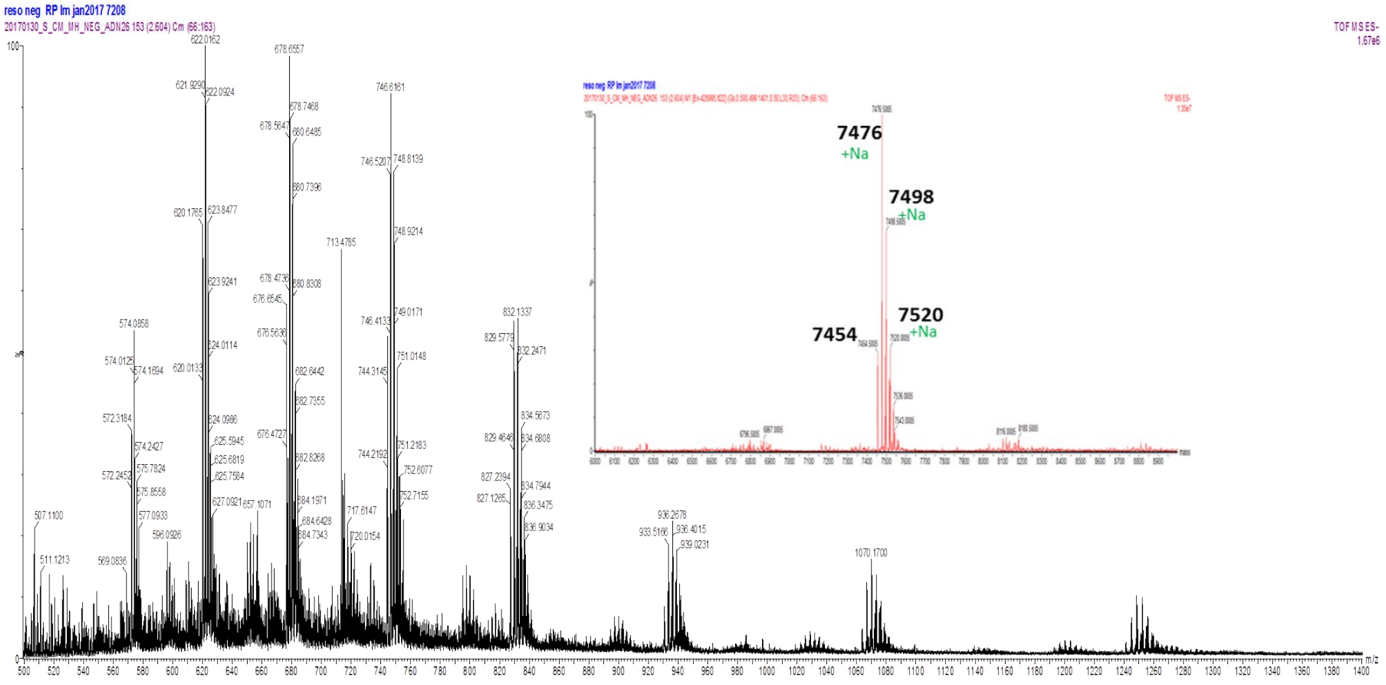


Figure S14. ESI$^-$ MS analysis of template **T1** in (AcN/H_2O/TEA 50/50/1).

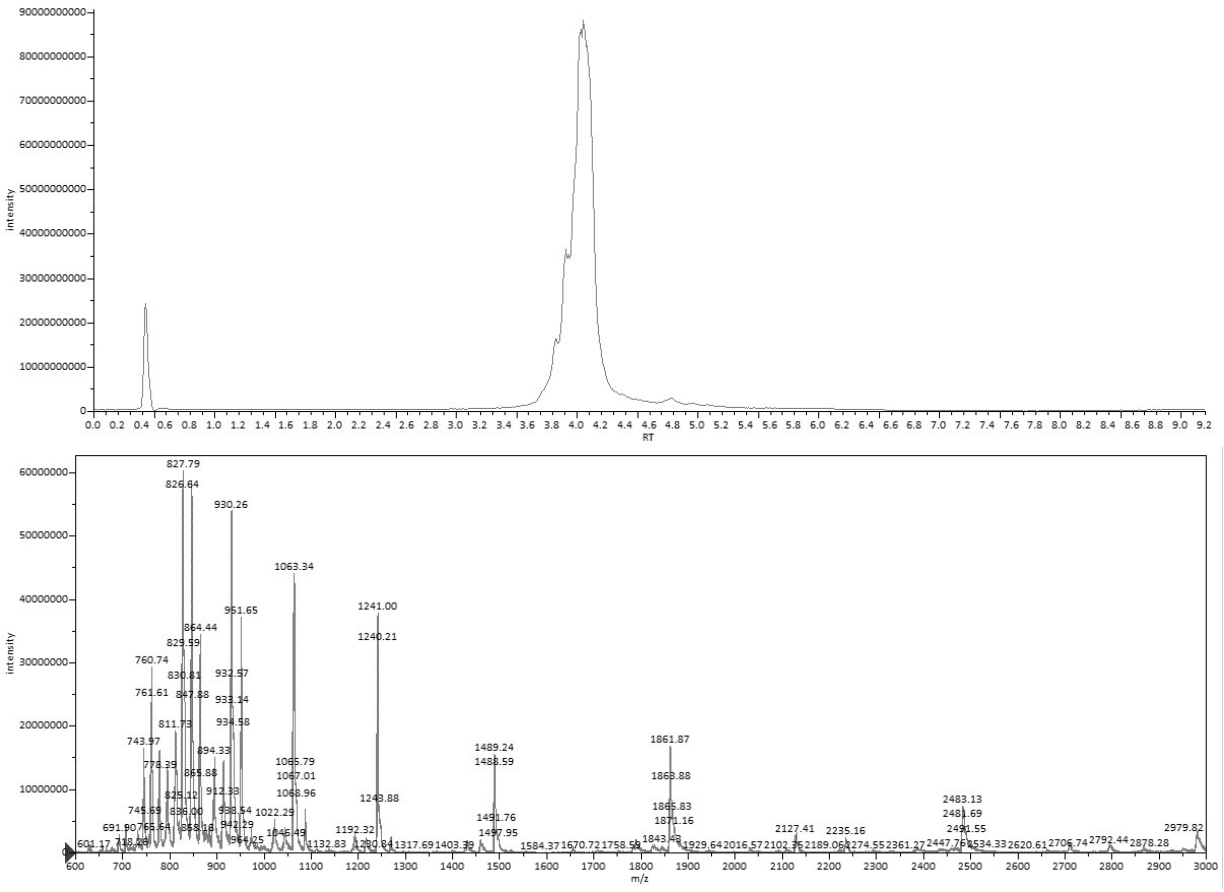


Figure S15. LC/MS analysis of template **T1**.

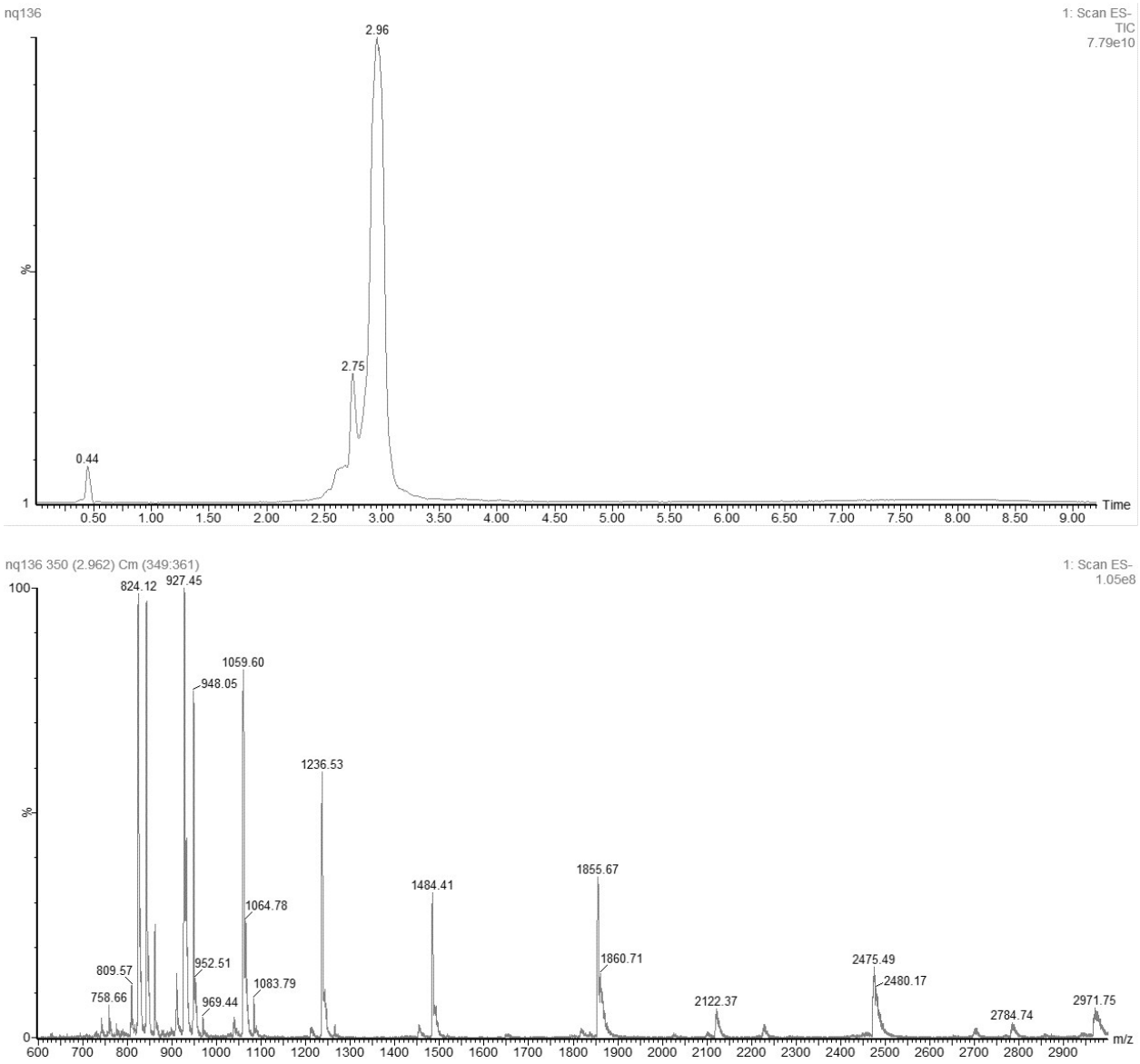


Figure S16. LC-MS analysis of template **T6**.

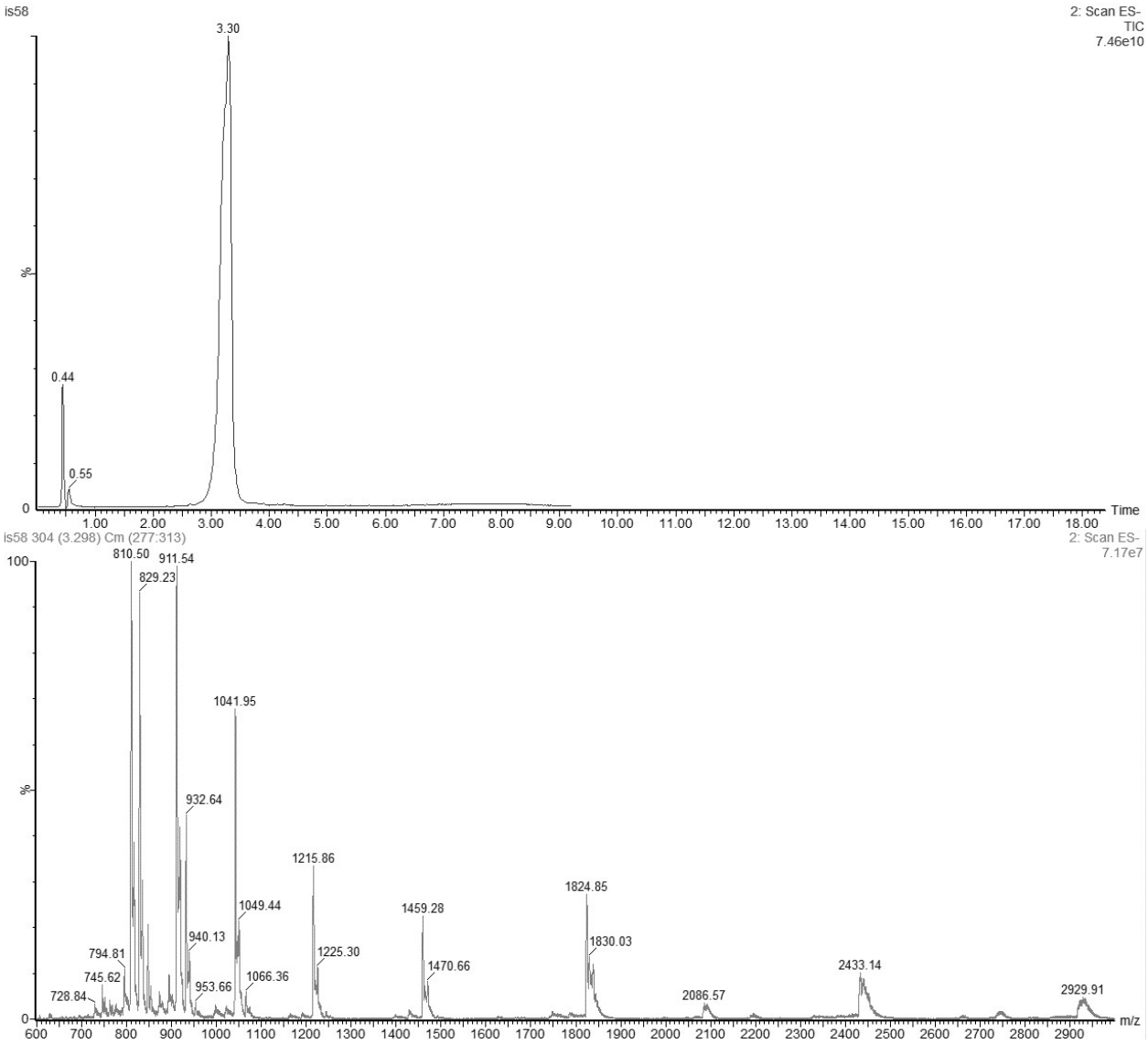


Figure S17. LC-MS analysis of template **T7**.

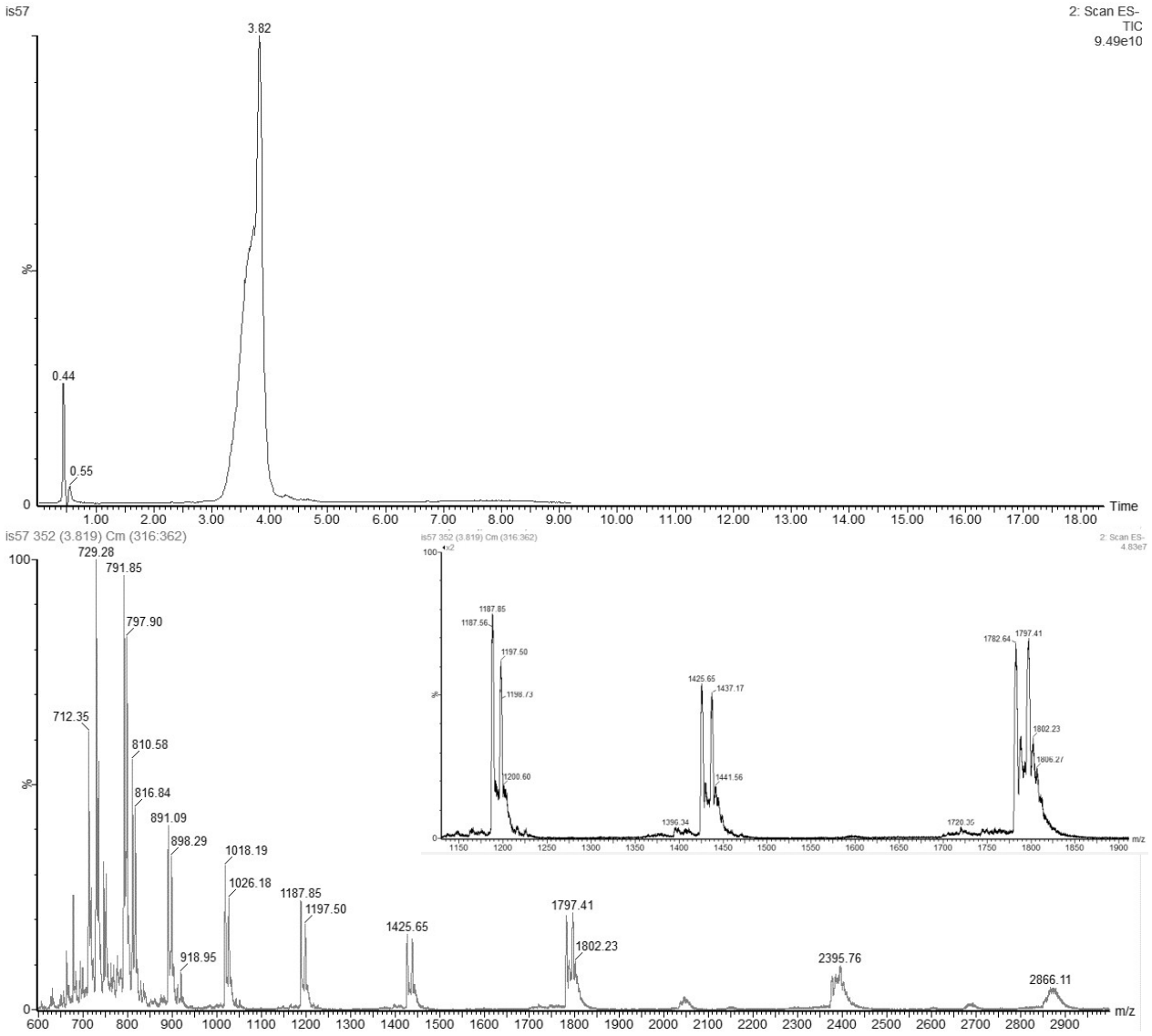


Figure S18. LC-MS analysis of template **T8**.

Table S1. Summary of the characterisation of the modified oligonucleotide templates.

Entry	Sequence (5'-->3')	m/z calc.	m/z found. (UPLC-MS)
T1	GGAG**Im**GAGGCTATAGTGAGTCGTA	7452.3	7451.8
T6	GGA**Im**TGAGGCTATAGTGAGTCGTA	7427.3	7426.7
T7	GG**ImImIm**GAGGCTATAGTGAGTCGTA	7302.3	7303.4
T8	**ImImImImIm**GAGGCTATAGTGAGTCGTA	7136.2	7134.6