Electronic Supplementary Material (ESI) for ChemComm.

SUPPORTING INFORMATION

4'-Guanidinium-modified siRNA: A molecular tool to control RNAi activity through RISC priming and selective antisense strand loading

Ganesh N. Nawale,[a] Saeed Bahadorikhalili,[a] Pallabi Sengupta,[b] Sandeep Kadekar,[a] Subhrangsu Chatterjee,[b] and Oommen P. Varghese[a]*

*Corresponding author

[a]Translational Chemical Biology Laboratory, Department of Chemistry, Ångström Laboratory, Uppsala University, 751 21, Uppsala, Sweden.

[b]Biomolecular NMR and Drug Design Laboratory, Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VII M, Kolkata, India.

EXPERIMENTAL SECTION

General. All chemicals and dry solvents were purchased from Sigma–Aldrich (Sweden) and used without any further purification. Unmodified phosphoramidites (N^6-benzoyl-rA, N^2-isobutyl-rG, N^4-acetyl-rC, rU and dT) and solid supports (Universal UnyLinker and dT Support, 1000Å) were purchased from ChemGene Corporation (USA). Deblocking [3 % trichloroacetic acid in DCM (w/v)], activator [0.25M 5-(ethylthio)-1H-tetrazole in ACN], Cap A [2,6-lutidine/acetic anhydride/THF: 8/1/1 (v/v/v)], Cap B [N-methylimidazole/pyridine/THF: 8/1/1 (v/v/v)] and oxidizer [pyridine/iodine/water/THF: 90.54/9.05/0.41/0.43 (v/v/v/w)] reagents used for RNAs synthesis were purchased from Sigma-Aldrich (Sweden). Thin layer chromatography (TLC) was performed on silica gel plates pre-coated with fluorescent indicator with visualization by UV light or by dipping into a solution of 5 % (v/v) conc. H_2SO_4 in EtOH and heating. ^{1}H NMR (400 MHz), ^{13}C NMR (100.5 MHz), ^{19}F NMR (376.2 MHz), and ^{31}P NMR (161.8 MHz) were recorded on 400 MHz instruments (Jeol JNM-ECP Series FT NMR). The chemical shifts in parts per million (δ) are reported downfield from TMS (0 ppm), or referenced to methanol-d4 (3.31 ppm), DMSO-d6 (2.50 ppm) for ^{1}H NMR spectra, and $CDCl_3$ (77.2 ppm), DMSO-d6 (39.5 ppm) or methanol-d4 (49.1 ppm) for ^{13}C NMR spectra. Multiplicities of ^{1}H NMR spin couplings are reported as s for singlet, bs for broad singlet, d for the doublet, t for the triplet, q for the quartet, dd for a doublet of the doublet, or m for multiplet and overlapping spin systems. Values for apparent coupling constants (*J*) are reported in Hz. High-resolution mass spectra (HRMS) for nucleoside intermediates were obtained in positive ion electrospray ionization (ESI) mode. The oligos and the coupling yields of the modified building blocks were determined using the absorbance at 495 nm using the trityl monitor. The coupling yields were more than 95% for each of the modified building blocks. Mass spectra of oligonucleotides were obtained in negative ESI mode. UV melting experiments were carried out in UV-visible spectrophotometer (Lambda 35 UV/Vis spectrophotometer from PerkinElmer) with Peltier temperature controller. Human serum (male AB) was obtained from Sigma-Aldrich (Sweden). In HCT116 cell culture experiments, the media DMEM, RNAiMAX, penicillin and streptomycin used was from Thermo Fischer Scientific (Sweden). Fetal bovine serum (FBS) used was from Hyclone (Sweden).

4′-*C*-guanidinicarbohydrazidomethyl-5-methyl uridine modified amidite synthesis

N-(2-Cyanoethoxycarbonyloxy)succinimide ***(1).*** A solution of 2-cyanoethanol (0.28 g, 4 mmol) in anhydrous CH_3CN (20 mL) was stirred at room temperature under argon atmosphere. To the above solution, was added pyridine (0.48 mL, 6 mmol) and *N,N′*-disuccinimidyl carbonate (1.28 g, 5 mmol) and stirred for more 1 h until the suspension became clear. After an additional 6 h of stirring, the reaction mixture was concentrated under reduced pressure. Saturated $NaHCO_3$ solution (100 mL) and DCM (100 mL) was added to the reaction mixture and organic phase was separated. The combined organic layer was washed with brine (50 mL), and dried over anhydrous Na_2SO_4, concentrated under reduced pressure. Traces of pyridine were removed by co-evaporation with dry ACN. The white solid was dried overnight in vacuum to afford compound **1**. (84 % Yield). R_f = 0.51 (5 % MeOH in DCM); ^{1}H NMR ($CDCl_3$): δ 2.74 (t, *J* = 8.0 Hz, 2H), 2.85 (s, 4H), 4.41 (t, J = 7.5 Hz); ^{13}C NMR ($CDCl_3$): δ 17.3, 25.2, 64.5, 117.1, 150.7, 168.7; HRMS (ESI) calcd. for $C_8H_8N_2O_5$ $[M]^+$ 213.0511, found $[M]^+$ 213.0514.

N,N-Bis-CEOC-2-methyl-2-thiopseudourea ***(2).*** *S*-methylisothiourea hemisulfate (0.22 g, 1.6 mmol) was suspended in DCM (20 mL) and sat. $NaHCO_3$ (20 mL). Compound **1** (0.85 g, 4.0 mmol) was added, and the reaction mixture was stirred for 2 hours at 45 °C. The organic phase was separated from the aqueous phase. The aqueous phase was extracted with DCM (2 × 20 mL), and the combined organic phase was dried ($MgSO_4$), filtered, and evaporated. The crude product was purified by flash silica gel column chromatography (ethyl acetate/DCM, 95:5) to afford compound **2**. (47 % Yield). R_f = 0.56 (5 % MeOH in DCM); ^{1}H NMR ($CDCl_3$): δ 2.45 (s, 3H), 2.75 (t, *J* = 8.0 Hz, 4H), 4.33 (p, *J* = 8.0 Hz, 4H); ^{13}C NMR ($CDCl_3$): δ 13.7, 18.2, 59.7, 59.8, 116.7, 117.1, 159.2, 161.1, 173.1; HRMS (FAB) Calcd for $C_{10}H_{13}N_4O_4S$ $[M]^+$ 285.0656, found: 285.0659.

2-cyanoethyl (Z) - ((((2-cyanoethoxy) carbonyl) amino) (2- (hydrazinecarbonyl) hydrazinyl) methylene) carbamate ***(3).*** A solution of compound **2** (60 mg, 0.2 mmol) in MeOH (1 mL) was added dropwise to a stirring solution of carbohydrazide (CDH) (90 mg, 1 mmol) in MeOH (2 mL). The reaction mixture was refluxed for 12 h. After the reaction completion, the solvent was evaporated, and the solid was washed 3 times with

MeOH/ DCM mixture (10:90, v/v) to collect the desired product **3** which was dried in vacuum (92 % Yield). R_f = 0.37 (5 % MeOH in DCM); ^{1}H NMR (DMSO-d_6): δ 2.87 (t, J = 7.2 Hz, 4H), 4.43 (t, J = 7.2 Hz, 4H), 7.14 (s, 2H), 8.31 (s, 1H); ^{13}C NMR (DMSO-d_6): δ 22.8, 57.5, 118.1, 153.1, 158.4, 164.8; HRMS (ESI) calcd for $C_{10}H_{14}N_8O_5$ $[M]^+$ 326.2691, found $[M]^+$ 326.2688.

2-((2R,3S,4R,5R)-3-(benzyloxy)-2-((benzyloxy)methyl)-4-hydroxy-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)acetonitrile ***(5)***. Nucleoside **4** (470 mg, 1 mmol) was dissolved in 10 mL of MeOH, and then, 1 M solution of NaOMe in MeOH (1.5 mL, w/v) was added, and stirred at room temperature for 3 h. After the reaction completion, the solvent was partially evaporated under reduced pressure and extracted with DCM (3 x 100 mL). The organic layer was collected and dried over anhydrous Na_2SO_4 and concentrated, and purified by flash column chromatography (40 % ethyl acetate in hexane) to get compound **5**. (91 % Yield). R_f = 0.44 (5 % MeOH in DCM); ^{1}H NMR ($CDCl_3$): δ 1.49 (s, 3H), 2.62 (d, J = 17.8 Hz, 1H), 2.89 (d, J = 17.8 Hz, 1H), 3.60 (d, J = 9.6 Hz, 1H), 3.82 (d, J = 9.6 Hz, 1H), 4.33 (d, J = 4.0 Hz, 2H), 4.53 (d, J = 11.6 Hz, 1H), 4.56 (s, 2H), 4.64 (d, J = 12.1 Hz, 1H), 4.85 (d, J = 12.2 Hz, 1H), 5.88 (d, J = 3.3 Hz, 1H), 7.25-7.37 (m, 13H), 9.99 (s, 1H) ; ^{13}C NMR ($CDCl_3$): δ 12.2, 22.7, 73.9, 74.6, 76.8, 77.1, 90.8, 110.9, 117.0, 127.9, 127.3. 128.4, 128.5, 128.7, 128.8, 136.0, 136.9, 137.0, 151.2, 163.7; HRMS (ESI) calcd for $C_{26}H_{27}N_3O_6$ $[M]^+$ 477.5101, found $[M]^+$ 477.5097.

2-((2R,3S,4R,5R)-3-(benzyloxy)-2-((benzyloxy)methyl)-4-((tert-butyldimethylsilyl)oxy)-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)acetonitrile ***(6)***. Imidazole (340 mg, 5 mmol) and compound **5** (427 mg, 1 mmol) were dissolved in DMF (5 mL), and then TBDMS-Cl (452 mg, 3 mmol) was added. The reaction mixture was stirred at room temperature for 2 h. After the completion of reaction, the reaction mixture was diluted with DCM. Saturated $NaHCO_3$ solution (100 mL) was added to the reaction mixture and organic phase was separated. The aqueous phase was extracted with DCM (3 x 100 mL). The combined organic layer was washed with brine (50 mL), and dried over anhydrous Na_2SO_4, concentrated under reduced pressure and purified by column chromatography (50 %

ethyl acetate in hexane) to get compound **6**. (87 % Yield). R_f = 0.61 (5 % MeOH in DCM); ^{1}H NMR ($CDCl_3$): δ 0.05 (s, 1H), 0.08 (s, 3H), 0.88 (s, 9H), 1.51 (s, 3H), 2.79 (d, J = 15.5 Hz, 1H), 3.00 (d, J = 15.8 Hz, 1H), 3.67 (d, J = 10.0 Hz, 1H), 3.90 (d, J = 9.4 Hz, 1H), 4.15 (d, J = 3.9 Hz, 1H), 4.49 (d, J = 4.4 Hz, 2H), 4.51 (d, J = 7.9 Hz, 2H), 4.87 (d, J = 11.6 Hz, 1H), 5.93 (d, J = 4.0 Hz, 1H), 7.23-7.25 (m, 3H), 7.30-7.36 (m, 6H), 7.49-7.50 (m, 1H); ^{13}C NMR ($CDCl_3$): δ 4.8, 5.0, 12.0, 18.1, 22.8, 25.7, 72.2, 73.9, 74.1, 75.6, 77.6, 84.4, 90.1, 100.0, 110.9, 116.9, 122.9, 128.1, 128.3, 128.5, 128.6, 128.8, 136.0, 136.8, 137.2, 150.4, 165.5; HRMS (ESI) calcd for $C_{32}H_{41}N_3O_6Si$ $[M]^+$ 591.7712, found $[M]^+$ 591.7714.

2-cyanoethyl ((2Z,8Z)-1-((2R,3S,4R,5R)-3-(benzyloxy)-2-((benzyloxy)methyl)-4-((tert-butyldimethylsilyl)oxy)-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)-13-cyano-5,10-dioxo-11-oxa-3,4,6,7,9-pentaazatridec-2-en-8-ylidene)carbamate ***(7)***. The nucleoside **7** was synthesized in 2 steps from compound **6**. In a round bottom flask was added compound **6** (430 mg, 0.75 mmol) and DCM (15 mL) and stirred at -78 °C for 15 min. Then 1 M solution of DIBAL in hexane (3.6 mL) was added dropsies during 30 min and stirred for more 2 h. After the reaction completion, the reaction mixture was quenched with MeOH at -78 °C followed by water and allowed to warm to room temperature. Saturated $NaHCO_3$ solution (100 mL) was added to the reaction mixture and organic phase was separated. The aqueous phase was extracted with ethyl acetate (3 x 100 mL). The combined organic layer was washed with brine (50 mL), and dried over anhydrous Na_2SO_4, concentrated under reduced pressure to give **7a**. The crude product was directly used in the next step for the synthesis of **7**. The compound **7a** was dissolved in MeOH (10 mL) and compound **3** (720 mg, 2.2 mmol) was added. The mixture was stirred under reflux conditions for 5 h. After the TLC monitoring had shown the reaction completion, the reaction mixture was cooled to room temperature, washed with saturated $NaHCO_3$ solution and extracted with DCM. The combined organic layer was washed with brine (50 mL), and dried over anhydrous Na_2SO_4, concentrated under reduced pressure and purified using flash column chromatography (5 % MeOH in DCM) to get compound **7**. (74 % Yield for two steps) R_f = 0.45 (5 % MeOH in DCM); ^{1}H NMR ($CDCl_3$): δ 0.08 (s, 6H), 0.88 (s, 9H), 1.59 (s, 3H), 2.60 (t, J = 7.7 Hz, 4H), 2.74 (d, J = 3.9 Hz, 2H), 3.62 (d, J = 11.6 Hz, 1H),

3.79 (d, J = 8.3 Hz, 1H), 3.86 (t, J = 7.9 Hz, 4H), 4.45 (d, J = 4.2 Hz, 1H), 4.51 (d, J = 12.2 Hz, 2H), 4.54 (s, 2H), 4.63 (d, J = 12.3 Hz, 1H), 5.29 (s, 1H), 6.08 (d, J = 3.5 Hz, 1H), 7.28-7.39 (m, 13H); ^{13}C NMR ($CDCl_3$): δ 4.2, 12.2, 17.4, 18.6, 22.4, 26.6, 59.2, 72.2, 74., 74.6, 75.0, 77.3, 85.0, 90.1, 99.5, 111.7, 117.5, 127.9, 128.3, 128.4, 128.5, 128.7, 136.1, 136.8, 148.4, 150.4, 155.8, 158.3, 163.4, 169.9; HRMS (ESI) calcd for $C_{42}H_{54}N_{10}O_{11}Si$ $[M]^+$ 903.0257, found $[M]^+$ 903.0259.

2-cyanoethyl ((Z)-1-((2R,3S,4R,5R)-3-(benzyloxy)-2-((benzyloxy)methyl)-4-((tert-butyldimethylsilyl)oxy)-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)-13-cyano-5,10-dioxo-11-oxa-3,4,6,7,9-pentaazatridecan-8-ylidene)carbamate ***(8)***. Nucleoside **7** (650 mg, 0.7 mmol) was dissolved in MeOH (15 mL) and stirred under argon atmosphere at room temperature. Further, sodium borohydride (95 mg, 2.2 mmol) was added in 3 portions. The reaction mixture was stirred at room temperature for more 1 h until the reaction completion. Saturated $NaHCO_3$ solution (100 mL) was added to the reaction mixture and organic phase was separated. The aqueous phase was extracted with DCM (3 x 100 mL). The combined organic layer was washed with brine (50 mL), and dried over anhydrous Na_2SO_4, concentrated under reduced pressure and purified by flash column chromatography (5 % MeOH in DCM) to get compound **8**. (94 % Yield). R_f = 0.48 (5 % MeOH in DCM); ^{1}H NMR ($CDCl_3$): δ 0.04 (s, 1H), 0.07 (s, 3H), 0.96 (s, 9H), 1.49 (s, 3H), 2.55 (t, J = 7.5 Hz, 2H), 2.77 (t, J = 8.1 Hz, 4H), 3.01 (d, J = 11.1 Hz, 1H), 3.48 (s, 2H), 3.65 (d, J = 11.8 Hz, 1H), 3.76 (d, J = 12.0 Hz, 1H), 3.88 (t, J = 8.3 Hz, 2H), 4.15 (d, J = 3.1 Hz, 1H), 4.41 (t, J = 7.6 Hz, 4H), 4.60 (d, J = 8.2 Hz, 2H), 5.21 (s, 1H), 5.93 (d, J = 3.3 Hz, 1H), 7.22-7.37 (m, 11H), 7.49 (s, 1H); ^{13}C NMR ($CDCl_3$): δ 4.7, 4.9, 13.0, 23.2, 26.6, 46.7, 54.3, 72.9, 74.7, 75.3, 75.7, 78.1, 85.8, 90.6, 112.4, 117.3, 128.7, 129.1, 129.2, 129.3, 129.5, 129.6, 136.9, 137.6, 139.2, 150.5, 155.2, 162.2, 163.3, 170.7; HRMS (ESI) calcd for $C_{42}H_{56}N_{10}O_{11}Si$ $[M]^+$ 904.3899, found $[M+H]^+$ 904.3902.

2-cyanoethyl ((Z)-3-(2-((2R,3S,4R,5R)-3-(benzyloxy)-2-((benzyloxy)methyl)-4-((tert-butyldimethylsilyl)oxy)-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)ethyl)-13-cyano-1,1,1-trifluoro-2,5,10-trioxo-11-oxa-3,4,6,7,9-pentaazatridecan-8-ylidene)carbamate ***(9).*** Compound **8** (610 mg, 0.67 mmol) was dissolved in THF and then ethyl trifluoroacetate (510 mg, 3.5 mmol) was added. The reaction mixture was stirred at room temperature for 4 h. After the reaction completion, saturated $NaHCO_3$ solution (100 mL) was added to the reaction mixture and organic phase was separated. The aqueous phase was extracted with DCM (3 x 100 mL). The combined organic layer was washed with brine (50 mL), and dried over anhydrous Na_2SO_4, concentrated under reduced pressure and purified by flash column chromatography (50 % ethyl acetate in hexane) to get compound **9.** (96 % Yield). R_f = 0.52 (5 % MeOH in DCM); ^{1}H NMR ($CDCl_3$): δ 0.16 (s, 6H), 0.63 (s, 9H), 1.66 (s, 3H), 2.64 (t, J = 7.4 Hz, 4H), 2.72 (t, J = 11.6 Hz, 2H), 3.44 (d, J = 9.8 Hz, 1H), 3.49 (t, J = 11.1 Hz, 2H), 3.52 (d, J = 8.9 Hz, 1H), 3.59 (d, J = 8.1 Hz, 1H), 4.35 (t, J = 7.5 Hz, 4H), 4.72 (d, J = 3.6 Hz, 1H), 5.73 (d, J = 7.7 Hz, 1H), 7.07-7.38 (m, 13H), 7.77 (s, 1H); ^{13}C NMR ($CDCl_3$): δ ;4.7, 4.9, 13.0, 20.3, 21.7, 23.2, 26.6, 46.7, 54.3, 72.9, 74.7, 75.3, 75.7, 78.1, 85.8, 90.6, 112.4, 112.6 (d, J = 161 Hz), 117.3, 128.7, 129.1, 129.2, 129.3, 129.5, 129.6, 136.9, 137.6, 139.2, 150.5, 155.2, 162.7, 163.2, 170.7; ^{19}F NMR ($CDCl_3$): δ -75.9; HRMS (ESI) calcd for $C_{44}H_{55}F_3N_{10}O_{12}Si$ $[M]^+$ 1000.3722, found $[M]^+$ 1000.3719.

2-cyanoethyl ((Z)-3-(2-((2R,3S,4R,5R)-4-((tert-butyldimethylsilyl)oxy)-3-hydroxy-2-(hydroxymethyl)-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)ethyl)-13-cyano-1,1,1-trifluoro-2,5,10-trioxo-11-oxa-3,4,6,7,9-pentaazatridecan-8-ylidene)carbamate ***(10).*** Nucleoside **9** (820 mg, 0.8 mmol) was dissolved in EtOH (15 mL) and $Pd(OH)_2$ supported on activated charcoal (120 mg) was added. Furhter, ammonium formate (250 mg, 4 mmol) was added, and the reaction mixture was heated at 70 °C for 48 h. The reaction mixture was filtered through celite pad and washed with EtOH (50 mL). The EtOH layer was evaporated and purified by flash column chromatography (5 % MeOH in DCM) to get compound **10**. (73 % Yield). R_f = 0.31 (5 % MeOH in DCM); ^{1}H NMR (DMSO-d_6): δ -0.06 (s, 3H), -0.01 (s, 3H), 0.78 (s, 9H), 1.19 (t, J = 7.1 Hz, 2H), 1.77 (s, 3H), 2.68 (t, J = 7.6 Hz, 4H), 3.74 (d, J = 9.8 Hz, 1H), 4.01 (d, J = 11.6 Hz, 2H), 4.21 (d, J = 8.5 Hz, 2H), 4.56 (t, J = 7.6 Hz,

4H), 4.75 (d, J = 4.2 Hz, 1H), 5.85 (d, J = 7.2 Hz, 1H), 8.94 (s, 1H); ^{13}C NMR (DMSO-d_6): δ -5.1, -4.7, 12.2, 18.6, 22.8, 25.7, 25.8, 27.2, 35.5, 43.6, 55.2, 72.2, 74.0, 74.2, 75.6, 84.4, 89.9, 111.5 (q, J = 158 Hz), 118.8, 136.1, 149.2, 149.4, 149.6, 149.8, 150.4, 155.4, 162.9 (q, J = 22 Hz), 170.4; ^{19}F NMR ($CDCl_3$): δ -74.3; HRMS (ESI) calcd for $C_{30}H_{43}F_3N_{10}O_{12}Si$ $[M]^+$ 820.2783, found $[M]^+$ 820.2785.

2-cyanoethyl ((Z)-3-(2-((2R,3S,4R,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((tert-butyldimethylsilyl)oxy)-3-hydroxy-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)ethyl)-13-cyano-1,1,1-trifluoro-2,5,10-trioxo-11-oxa-3,4,6,7,9-pentaazatridecan-8-ylidene)carbamate ***(11)***. Compound **10** (470 mg, 0.56 mmol) rendered anhydrous by repeated co-evaporation with dry pyridine and finally dissolved in the same solvent (10 mL) and DMAP (140 mg, 1.1 mmol) was added to the above solution. The reaction mixture was stirred at 45 °C, and then DMTr-Cl (470 mg, 1.3 mmol) was added to the above solution in 3 portions during 1 h. After completion of the reaction, reaction mixture was cooled to room temperature and was diluted with DCM followed by extracted with DCM (3 x 100 mL). The combined organic layer was dried over anhydrous Na_2SO_4, concentrated under reduced pressure and purified by flash column chromatography (5 % MeOH in DCM + 2 % Et_3N) to get trytilated compound **11.** (93 % Yield). R_f = 0.49 (5 % MeOH in DCM + 2 % Et_3N); ^{1}H NMR ($CDCl_3$): δ 0.07 (s, 3H), 0.08 (s, 3H), 0.88 (s, 9H), 1.61 (t, J = 6.0 Hz, 2H), 1.66 (d, J = 8.0 Hz, 2H), 2.10 (s, 3H), 2.60 (t, J = 6.0 Hz, 4H), 2.74 (d, J = 8.0 Hz, 2H), 3.17 (t, J = 8.0 Hz, 2H), 3.36 (s, 6H), 3.62 (d, J = 12.0 Hz, 1H), 3.76 (t, J = 6.0 Hz, 4H), 3.79 (d, J = 8.0 Hz, 1H), 4.45 (d, J = 8.0 Hz, 1H), 4.49 (d, J = 8.0 Hz, 1H), 4.53 (s, 2H), 4.62 (d, J = 8.0 Hz, 1H), 5.29 (s, 1H), 5.49 (t, J = 4.0 Hz, 1H), 6.08 (d, J = 4.0 Hz, 1H), 6.81 (dd, J_1 = 8.0 Hz, J_2 = 1.0 Hz, 4H), 7.14 (dd, J_1 = 8.0 Hz, J_2 = 1.0 Hz, 4H), 7.22-7.25 (m, 6H), 7.25-7.26 (m, 1H), 7.28-7.39 (m, 6H); ^{13}C NMR ($CDCl_3$): δ 3.2, 3.9, 12.2, 16.9, 20.8, 22.4, 23.1, 25.2, 55.5, 56.2, 58.0, 72.2, 74.0, 74.6, 75.0, 77.3, 85.0, 99.9, 101.2, 109.3, 111.7 (q, J – 161 Hz), 112.8, 113.5, 113.6, 116.5, 122.9, 128.3, 128.5, 128.7, 128.8, 136.0, 136.8, 146.6, 150.2, 150.7, 153.9, 159.5, 169.9, 170.8, 176.3; ^{19}F NMR ($CDCl_3$): δ -75.7; HRMS (ESI) calcd for $C_{52}H_{62}F_3N_9O_{14}Si$ $[M]^+$ 1121.4138, found $[M]^+$ 1121.4140.

2-cyanoethyl ((Z)-3-(2-((2R,3S,4R,5R)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((tert-butyldimethylsilyl)oxy)-3-(((2-cyanoethoxy)(diisopropylamino)phosphanyl)oxy)-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)ethyl)-13-cyano-1,1,1-trifluoro-2,5,10-trioxo-11-oxa-3,4,6,7,9-pentaazatridecan-8-ylidene)carbamate *(**12**).* The tritylated nucleoside **11** (580 mg, 0.5 mmol) was dissolved in DCM (10 mL) followed by addition of DIPEA (325 mg, 2.5 mmol) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (475 mg, 2 mmol). The reaction mixture was stirred at room temperature for 3 h. Further, MeOH (0.5 mL) was added to the reaction mixture and stirred for 5 min followed by DCM (25 mL) and saturated $NaHCO_3$ (25 mL) were added. The organic layer was extracted, washed with brine (2 x 25 mL), dried over Na_2SO_4, and filtered. The product was co-evaporated with acetonitrile and further purified by flash silica column chromatography (100 % DCM + 2 % Et_3N) to get the product **21**. (76 % Yield). R_f = 0.83 (5 % MeOH in DCM + 2 % Et_3N); ^{31}P NMR ($CDCl_3$) δ 151.14, 151.17; ^{19}F NMR ($CDCl_3$) δ −77.05, −77.07; HRMS (ESI) calcd for $C_{60}H_{78}F_3N_{12}O_{15}PSi$ $[M+H]^+$ 1322.5169, found $[M+H]^+$ 1322.5172.

RNA synthesis

All siRNA sequences were synthesized using an automated solid-phase synthesizer (H8, K&A Synthesizer) with 2′-O-TBDMS protected A, C, G and U monomers and dT monomer employing standard synthesis cycle. Universal UnyLinker or dT support (1000Å) was used for synthesis. All amidites were dissolved in ACN with 0.06 M concentration. The coupling time used for unmodified phosphoramidites was 6.5 minutes, and for the modified phosphoramidites was 15 minutes. The cleavage from the solid support and deprotection of guanidinium and phosphate protecting groups were carried out using 50 % piperidine in water at room temperature for 24 h. Further base deprotection was carried out with the treatment of 1:1 mixture of aq. methylamine and aq. NH_3 (v/v, AMA) for 18 min at 65 °C. The 2′-*O*-tert-butyldimethylsilyl (TBDMS) group were removed using $Et_3N.3HF$ in DMSO at 65 ^oC for 2.5 h and purified by 20 % denaturing (7 M Urea) PAGE. The RNA samples were recovered with TEN buffer (1 mM EDTA, 10 mM Tris, 300 mM NaCl, pH 8) and desalted using Sep-Pak (WAT020515,

Waters) column. The pure RNA pellet was dissolved in water, and the concentration was measured at 260 nm in UV−VIS spectrophotometer.

Thermal stability of siRNA

Melting temperature analysis was carried out by measuring the change in absorbance at 260 nm with increase in temperature using a UV spectrophotometer coupled with a Peltier controller. Duplexes were prepared using equimolar concentration of complementary strands in water (1 μM siRNA, 10 mM phosphate and 50 mM NaCl). The solution containing passenger strand and guide strands was heated at 95 °C for 2 min, then gradually cooled to room temperature over a period of 3 h, and stored at 4 °C. Before the measurements, the samples were equilibrated at 25 °C for 5 min and then heated to 95 °C in an increment of 1 °C/min. The change in absorbance against an increase in temperature was recorded, and T_m was calculated utilizing the Boltzmann fitting or first-order derivative in Origin 2015. All experiments were triplicated, and the T_m values reported are an average of 3 independent measurements.

Serum stability of siRNA

Passenger and guide strands were mixed at an equimolar ratio in nuclease-free water. The solutions were heated at 95 °C for 2 min and then gradually cooled to room temperature. The annealed siRNA duplexes (30 μM) were incubated to 70 % human serum (from male AB clotted whole blood) with a final volume of 80 μL. The aliquots of 10 μL were withdrawn after 1, 2, 4 and 8 h; and then quenched with 5 μL of 6X stop dye (10 mM Tris-HCl, pH 7.6, 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, 60 % glycerol and 60 mM EDTA) and stored at −20 ^{o}C till loading on gel. The samples were analyzed by 15 % native PAGE (0.7 mm thickness, 15-20 W) with 1X TBE (89 mM of each tris and boric acid, 2 mM of EDTA, pH 8.3) as running buffer. Gels were stained with SYBR® Green (ThermoFischer scientific, Sweden) stain for 15 minutes and visualized by using UV.

***In-vitro* transfection**

HCT116 (human colon cancer) cell line were obtained from American Type Culture Collection, ATCC (Manassas, VA, USA). A day prior to transfection, the cells were seeded at a density of 50,000 cells in a 24-well cell culture plate to achieve ~70-80 %

confluence at the time of transfection. DMEM consisting of 10 % of FBS, and 1 % antibiotics (PeSt) at 37 °C and 5 % CO_2 was used as a medium. On the day of transfection, the medium was replaced with a fresh complete, DMEM. The siRNAs (unmodified and modified) were transfected with different concentrations of siRNA employing RNAiMAX reagent (Thermo Fischer Scientific, Sweden) for siRNAs according to the manufacturers' protocols. Briefly, the siRNA was mixed with RNAiMAX reagent and incubated for 10 min. This complex was subsequently added to the cells to be transfected. Cells were also transfected with negative control siRNA (scrambled sequence). Each transfection was performed in triplicate. Post-transfection, cells were incubated for 24/48/96 h.

Total RNA samples

After incubation as stated above, total RNA was isolated from cells by adding 350 µl of lysis buffer (Qiagen, Germany), followed by homogenization of cell lysates. RNA was extracted from cell lysates by RNeasy Mini Kit (Qiagen, Germany). The NanoDrop 2000 (Thermo Fischer Scientific, Sweden) was used to determine RNA concentrations, with resulting OD 260/280 ratios between 1.85-2.03.

Real-time RT-PCR

1 microgram of the total RNA was used to make the cDNA. The cDNA was prepared using High Capacity RNA to cDNA kit according to manufacturer's protocol (Applied Biosystems, USA) and qRT-PCR was performed with 1000 ng of cDNA with TaqMan® Fast Universal PCR Master Mix (2X). The real-time PCR reactions were carried out with 10 µl of 2x TaqMan® Universal PCR Master Mix, no AmpErase® UNG (Applied Biosystems, USA), 5 µl diluted cDNA, and 1 µl of TaqMan gene-specific assay mix in a 20 µl final reaction volume. Reference gene, β-actin (*ACTB*) (Applied Biosystems, USA) was selected as a control for normalization of real-time PCR data. The amplification was carried out using the Biorad CFX connect system using a 40-cycle program. The CFX manager software automatically calculates the raw Ct (cycle threshold) values. Data from samples with a Ct value equal to or below 30 were further analyzed. Samples were normalized relative to endogenous control and differences in cycle number thresholds were calculated using comparative quantitation $2^{-\Delta\Delta CT}$ method (also called the ΔΔCT

method), which is commonly used for analyzing siRNA induced gene knockdown efficiency.

The formulas used to calculate gene knockdown were as follows:

First, the ΔCT was calculated as the mean cycle threshold for the target gene minus the mean cycle thresholds for the endogenous controls *ACTB*, each performed in triplicates:

ΔCT = CT (target gene) – CT (endogenous control). Secondly, the ΔΔCT was calculated as the ΔCT of the target minus the ΔCT of negative control (NC): ΔΔCT = ΔCT (target) – ΔCT (NC). Thereafter, the percentage of knockdown of the target gene was calculated as:

Fold change = 2−ΔΔCT, then percentage of Knockdown: = 100 * (1-fold change)

Stem-loop real-time PCR

After transfection, RNA was isolated (24 h post transfection) using the mirvana miRNA isolation kit (Thermo Fischer Scientific, Sweden) by following the manufacturer's protocol. Stem-loop primers were used to make cDNA by using 10 ng of RNA. Stem-loop primers and qPCR TaqMan primers specific for each strand of the *STAT3* siRNA were custom ordered (Thermo Fischer Scientific, Sweden). The *U6* RNA was used as endogenous control, the stem-loop primers and real-time PCR Taqman primers for *U6* were obtained from Thermo Fischer Scientific, Sweden. The cDNA synthesis was done in accordance with the manufacturers' protocol. Real-time PCR was done in the same manner as described in the previous section.

***In silico* structure preparation of siRNA duplexes**

We constructed three-dimensional structures of siRNA1 and siRNA6 using Nucleic Acid Builder (NAB) web server and further introduced guanidinium modifications in Maestro. We refined these models using the Protein Preparation Wizard (PPW) in Maestro, followed by restrained minimization using the IMPREF utility, which applies the constraint to converge the non-hydrogen atoms of the system to the root-mean-square deviation (RMSD) of 0.30 Å using OPLS2005 (optimized potentials for liquid simulations) force field (PMID: 16211539).[63] It corrects the disallowed torsion angles in the system by eliminating unfavorable atom-atom contacts (PMID: 23579614),[64-65] but ignores the long-range non-bonded interactions. Therefore, we further minimized the

predicted structures with the same force field (OPLS2005) and implicit solvation. We used the Powell-Reeves Conjugate Gradient minimizer (PRCG) with a convergence threshold of 0.05 kJ/Å-mol and 5000 number of iterations. The cut-off distances are, by default, 7 Å for Van der Waals, 12 Å for electrostatics, and 4 Å for hydrogen bonds.

Molecular docking and minimization

We performed docking between RNase A-siRNA1, RNase A-siRNA6, RISC-siRNA1 and RISC-siRNA6 using Z-DOCK 2.3.2. Z-DOCK employs the Fast Fourier transform (FFT) algorithm and utilizes a scoring function that combines the pairwise shape complementarity with short- and long-range electrostatic and Van der Waal energies to enable highly efficient docking. This increases the number of highly accurate conformations in the initial selection of the poses while reducing the number of false positives (*e.g.*, geometrically distant complexes with good scores) and the poses that require computationally expensive electrostatic calculations. Furthermore, we applied biasing constraints over the catalytic core of RNase A (His119, Cys41, and His12) and the PAZ domain of Argonaute protein. The constraint sets an additional filter, which improves the docking simulations by pruning the search space and yields more accurate docking poses by scanning a smaller set of inter-protein contacts (PMID: 25722738).[66] The constrained docking simulation gives 10 most promising candidate models. We established the docking benchmarks by superimposing the docked poses from Z-DOCK and clustered the similar poses, which are reproduced independently by the docking protocols. We minimized the complexes in Maestro using the aforementioned protocol.

Molecular dynamics (MD) simulation

We performed MD simulation over minimized siRNA duplexes using the Desmond module of Maestro. We used sodium ions to neutralize the system and submerged the system into TIP3P water model with box boundaries 10 Å away from the outermost complex atoms along x-, y-, and z-directions. This gives a more intuitively realistic picture of intermolecular interactions that retains specific solvent interactions as occurs in the cellular system. The solvated complexes undergo two-step minimization in vacuum to 0.001 kcal/mol/Å of RMS gradient of potential energy to eliminate the steric

clashes, if any. After minimization, we carried out MD simulation at a time step of 1 fs and gradually increased the temperature of the system to 300 K over 50 ps simulation under NVT conditions. Next, we carried out the simulations for another 50 ps under NPT conditions at 1 atm pressure. Then, we simulated the equilibrated complexes for the production dynamics for 100 ns under NVT conditions. Temperature and pressure are regulated using a Langevin thermostat-barostat. We applied the Particle Mesh Ewald method for long-range electrostatic interactions with a grid space of 0.1 nm and set the non-bonded cut-off at 12.0 Å. All the calculations are performed for 2 fs time window and coordinates are recorded at an interval of 10 ps. An ensemble of structures from last 10 ns of MD simulations for complexes is used for the evaluation and analyses of the binding contacts. The snapshots are extracted at an interval of 50 ps resulting in 10 conformations of the complexes for each 100 ns trajectory. The intramolecular contacts involving guanidine group and terminal bases are identified using a distance cut-off of 4.0 Å. Only those contacting residue pairs, which are found in more than 80 % of the complexes in last 10 ns of the MD trajectories are considered as the residue pairs having stable interactions.

Sugar puckering analyses by xDNA

Sugar puckering in ribose sugars is an index of understanding RNA conformations, which is also a key deterministic factor for siRNA loading into RISC complex. Therefore, we calculated the sugar puckering of ribose sugars of the simulated siRNA1 and siRNA6 molecules employing xDNA software. The calculations allow us to quantitatively estimate sugar endocyclic (v0 (C4'–O4'–C1'–C2'), v1 (O4'–C1'–C2'–C3'), v2 (C1'–C2'–C3'–C4'), v3 (C2'–C3'–C4'–O4'), and v4 (C3'–C4'–O4'–C1')) and exocyclic (δ (C5'–C4'–C3'–O3')) dihedrals and the modes of sugar puckering. With this information we calculated the Pseudorotation phase angle P using the following formula:

$$\tan P = \frac{(v1 + v4) - (v0 + v3)}{2 \times v2(sin36^0 + sin72^0)}$$

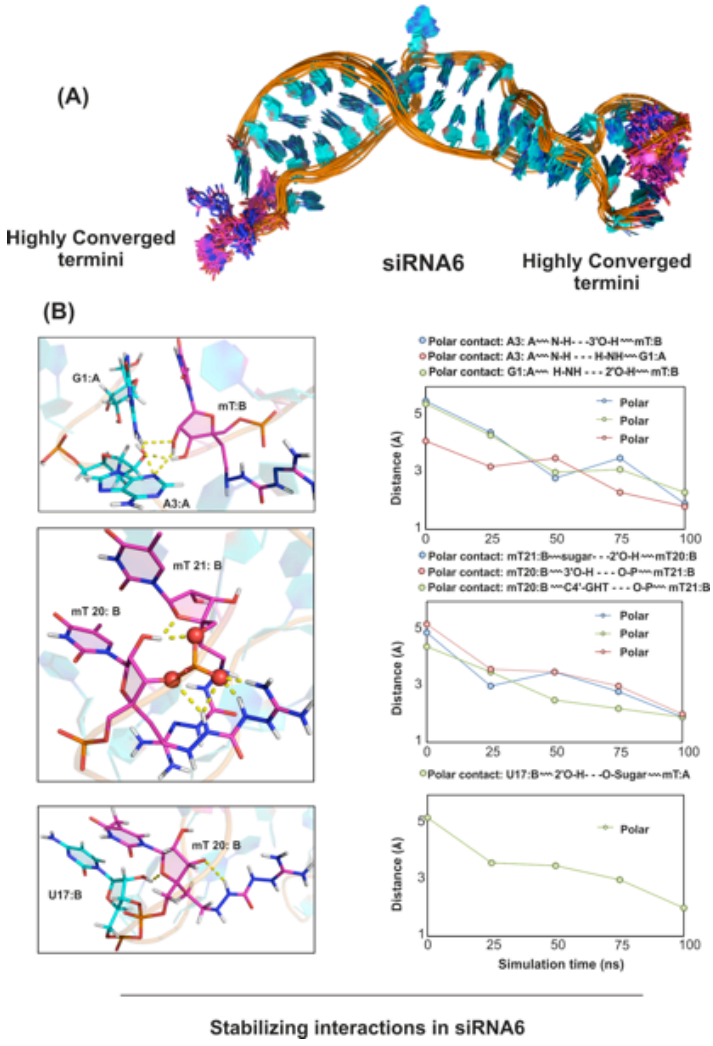
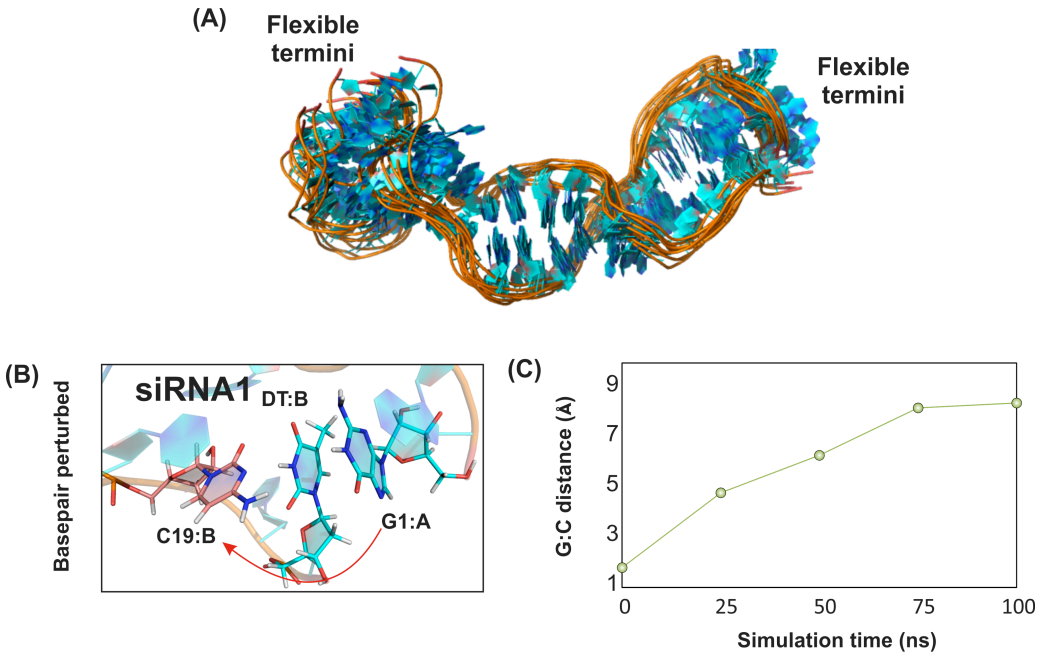


Fig. S1. The GMU modification at 3′-termini imparts thermodynamic stability to siRNA duplex. (A) The ensemble of the GMU modified **siRNA6** is shown over the last 10 ns simulation. (B) The residues that actively participate in stable non-covalent interactions in **siRNA6** are shown along with the distances (in angstroms) between interacting atoms at 25 ns interval over 100 ns of the production run. We observed that the siRNA models attained stable states over 100 ns simulation (Figure S1A). In **siRNA6**, dual GMU modifications at 3′-overhang amplified the network of electrostatic interactions with adjacent residues and built a closure at 3′-termini to protect proximal base pairs. The penultimate guanidinium linker was engaged in an H-bonded network with 2′-OH of adjacent uridine (U17 in chain B) sugar whereas its phosphate backbone was stabilized by an array of H-bonds by the terminal guanidinium chain (Figure S1B). Three Watson-Crick base pairs at the termini were distorted. The guanine (G1 in chain A) and adenine (A3 in chain A) residues were connected through amino-imino contacts and exhibited bifurcated hydrogen bonds with 2′-OH of terminal ribose sugar in the modified residue (Figure S1B). The network of coulombic interactions by dual guanidinium linkers preserved terminal base pairs, which might retard the unfolding of double-helical RNA structures.



(D)

	siRNA6	siRNA1
Total energy (E)	-1,12,144.079 kcal/mole	-12,004.09 kcal/mole
Potential energy (PE)	-1,35,975.178 kcal/mole	-36,575.78 kcal/mole
Degrees of freedom (df)	80,233	30,233
RMSD (Å)	2.53	4.9 (termini) 2.1 (central)

Figure S2. Thermodynamic stability of unmodified **siRNA1** duplex. (A) The ensemble of **siRNA1** duplexes are shown over the last 10 ns of 100 ns simulation. (B) The terminal thymidine in chain B enabled π-π stacking interaction between neighbouring guanines in chain A, which sequestered the cytosine residue, C19 in the complementary chain (C) Base pairing distance between G1 and C19 are shown over the 100 ns simulation, (D) The average RMSD values of 3′-termini in modified and unmodified siRNA duplexes across 100 ns, total energy and potential energy of the systems are demonstrated to analyse MD simulation trajectory. The modified siRNA structures (**siRNA6**) converged to steady RMSD values at a range of 1.5 – 2.5 Å (Figure S2D) unlike the unmodified siRNA1, which did not converge at terminal overhangs but showed steady RMSD (Root mean square deviation) values at the central portion of the double-helix (Figure S2A and S2D).

***Table S1*. Thermodynamic parameters calculated using the MM-PBSA method from 100 ns simulation data**

	siRNA1 (kcal.mol^{-1})	**siRNA6** (kcal.mol^{-1})
ΔE_{elec}	-352.7± 54.9	-576.9 ± 78.2
ΔE_{vdw}	-28.45 ± 3.69	-44.86 ± 4.28
ΔE_{mm} ($\Delta E_{elec} + \Delta E_{vdw}$)	-381.15	-621.76
ΔG_{PB}	398.2 ± 28.2	599.4 ± 51.2
ΔG_{np}	-20.2 ± 1.23	-38.9 ± 2.09
ΔG_{solv} ($\Delta G_{PB} + \Delta G_{np}$)	378	560.5
ΔG_{Total} ($\Delta E_{mm} + \Delta G_{solv}$)	-3.15	-61.26

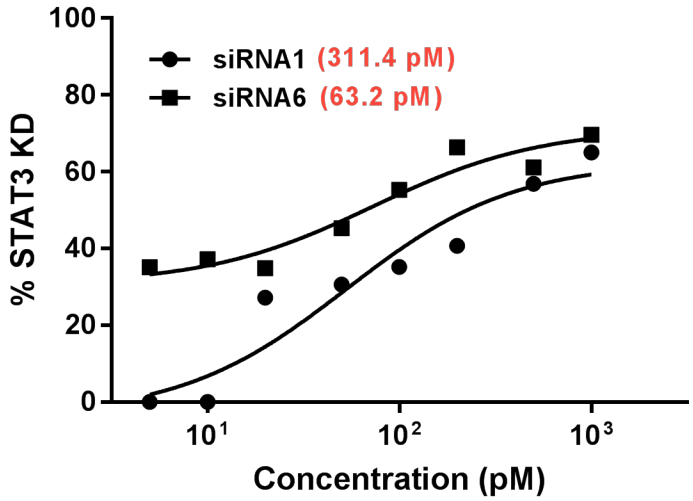


Figure S3. Concentration-dependent RNAi activity post 96 h of transfection. Experiment were performed in duplicate for two times. Analysis was performed using GraphPad Prism software. IC50 determination was performed using non-linear regression analysis (log [inhibitor] vs.normalized response).

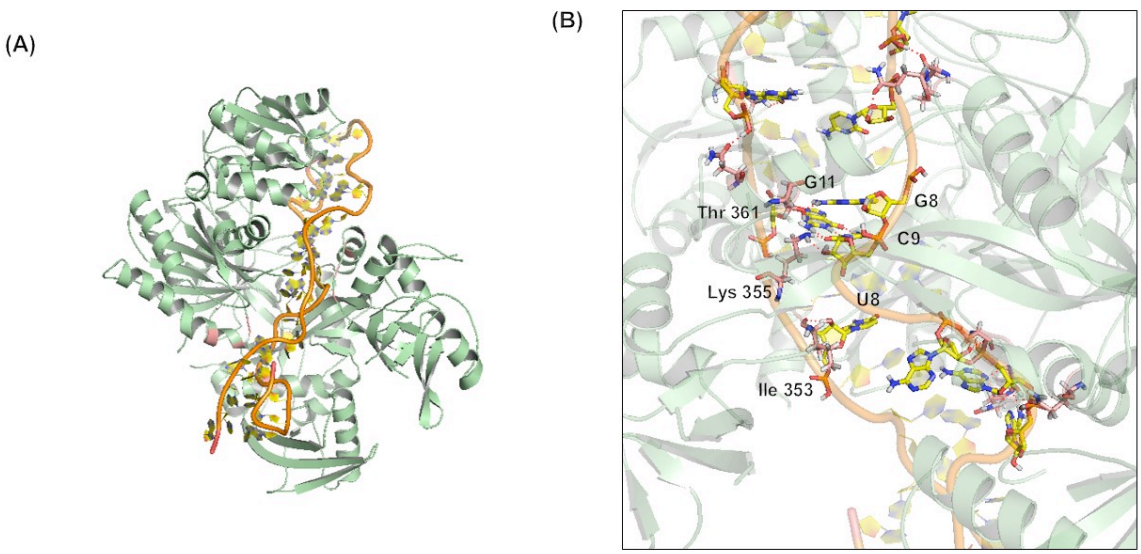


Figure S4. Interaction of RISC-siRNA1 complex A) **siRNA1** recruitment into the PAZ domain of RISC complex. B) Non-covalent interactions between **siRNA1** and PAZ domain.

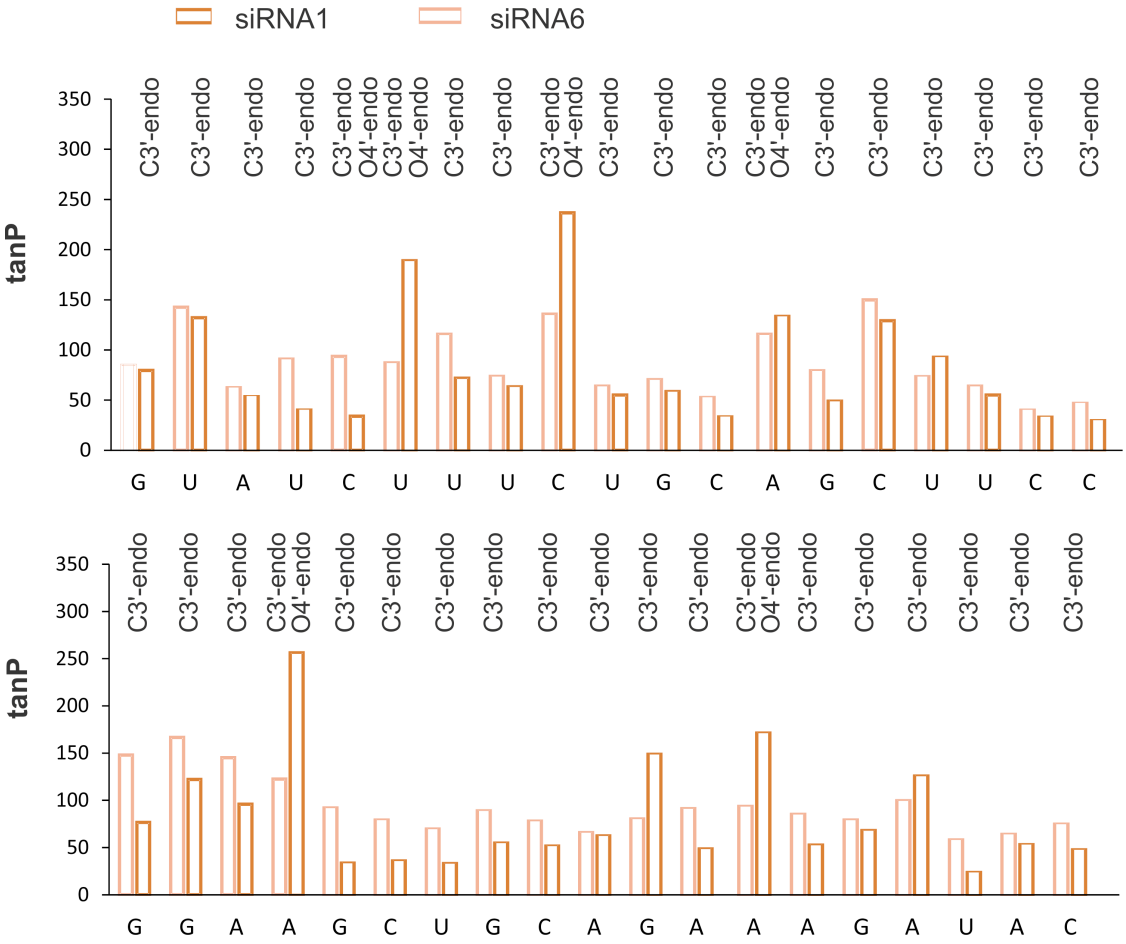
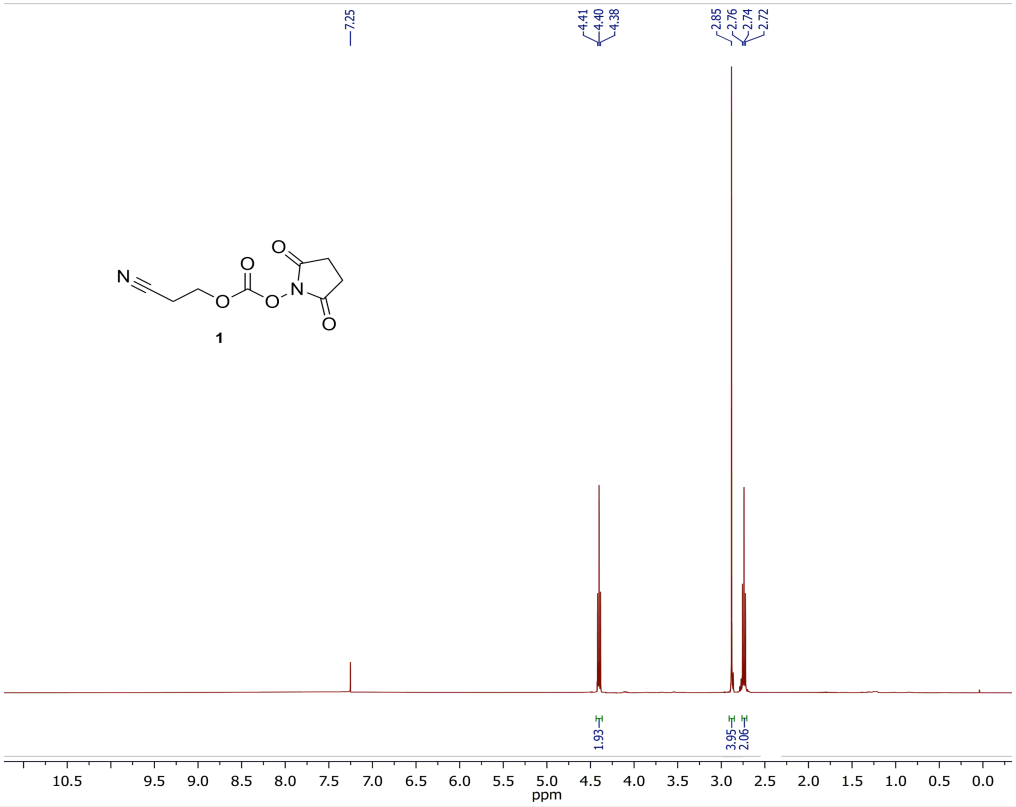
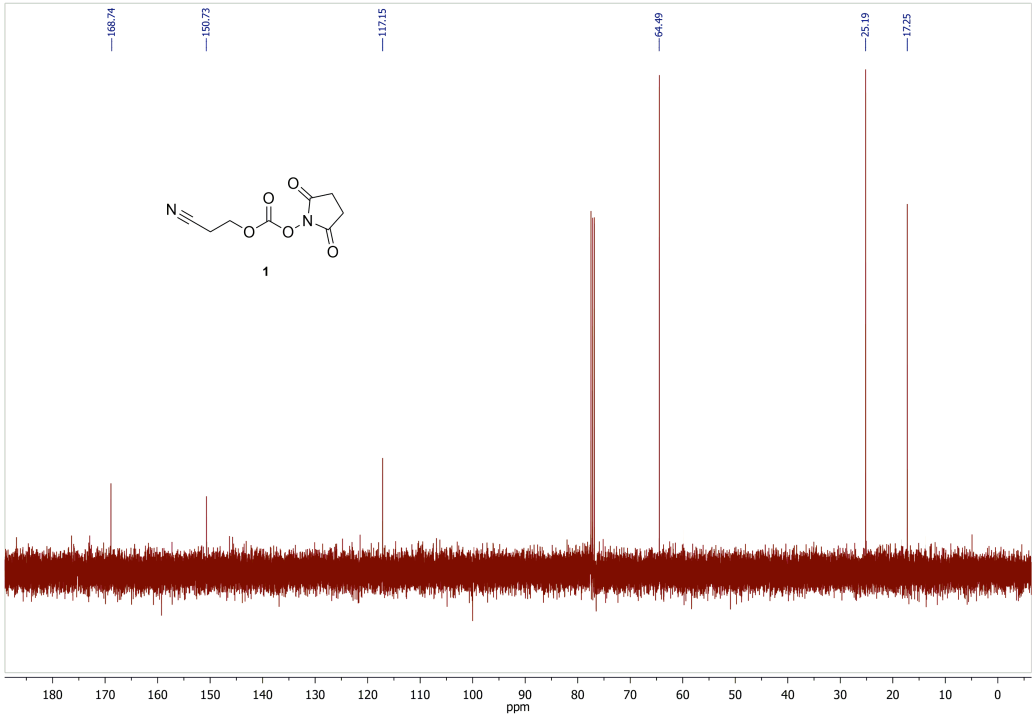


Figure S5: Sugar puckering analyses of **siRNA1** and **siRNA6**.

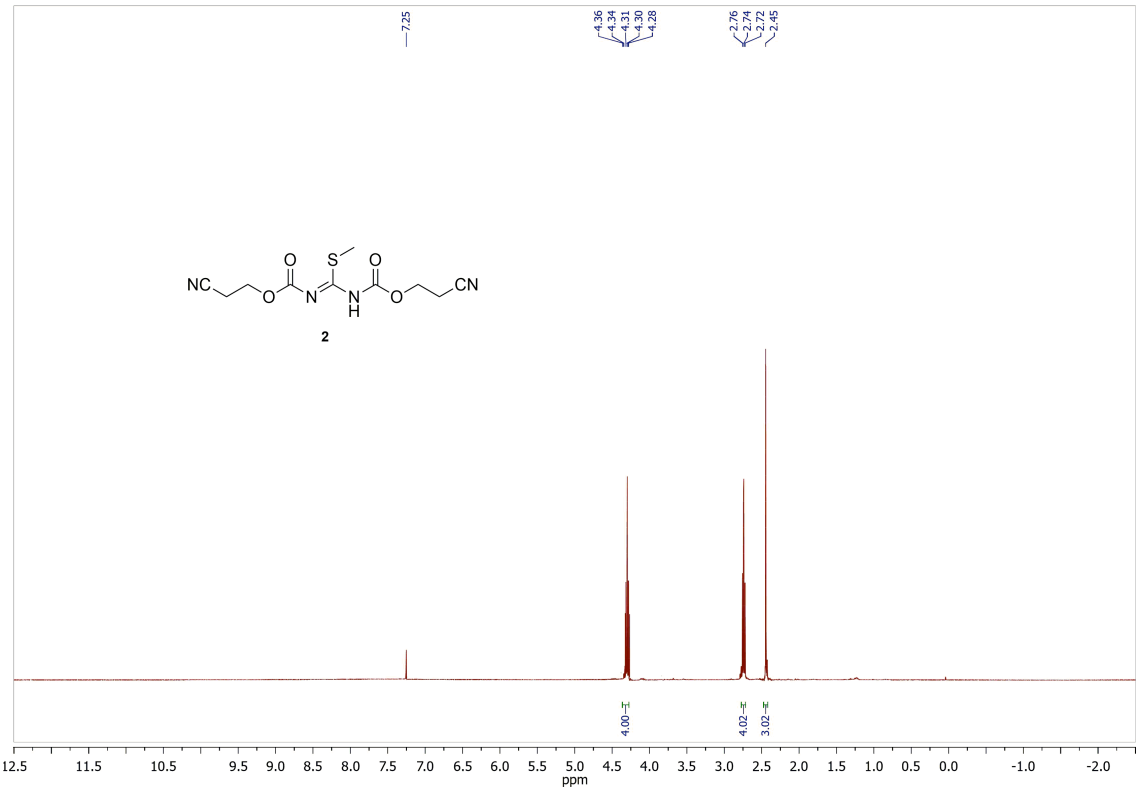
NMR spectra's for compounds

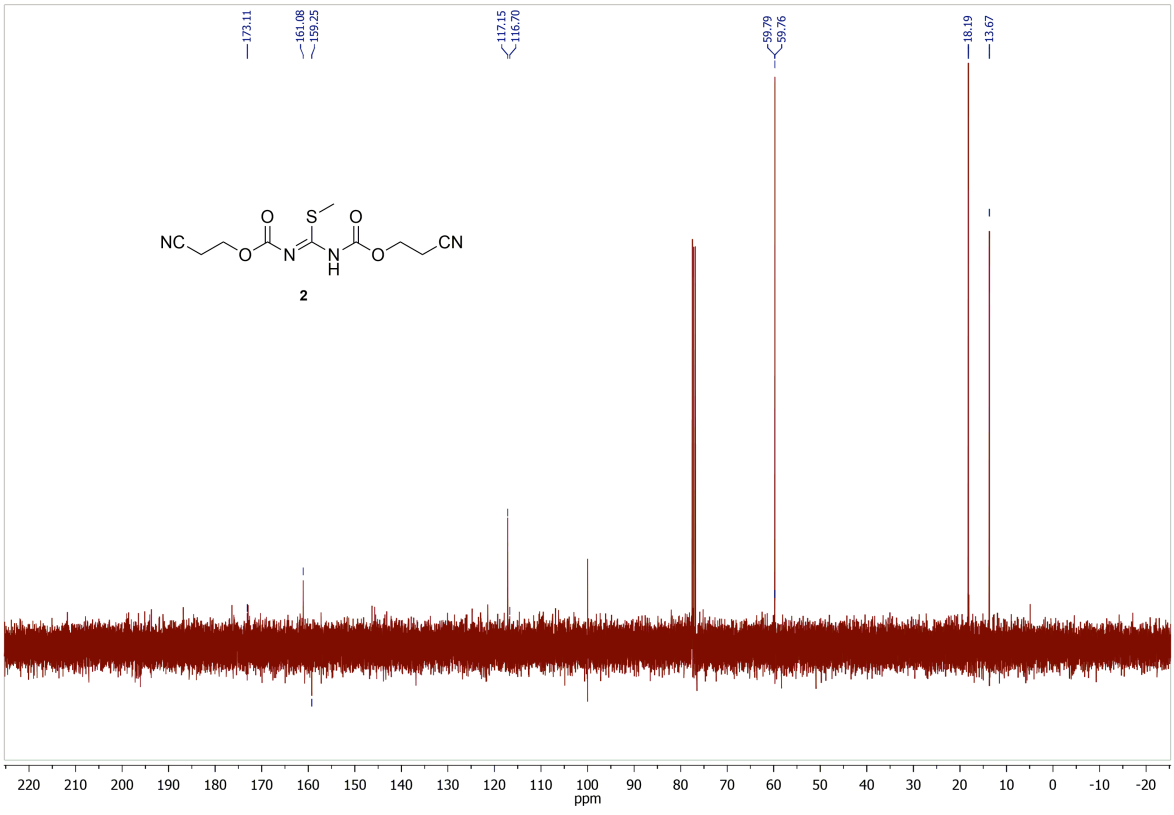
^{1}H and ^{13}C NMR for compound **1**



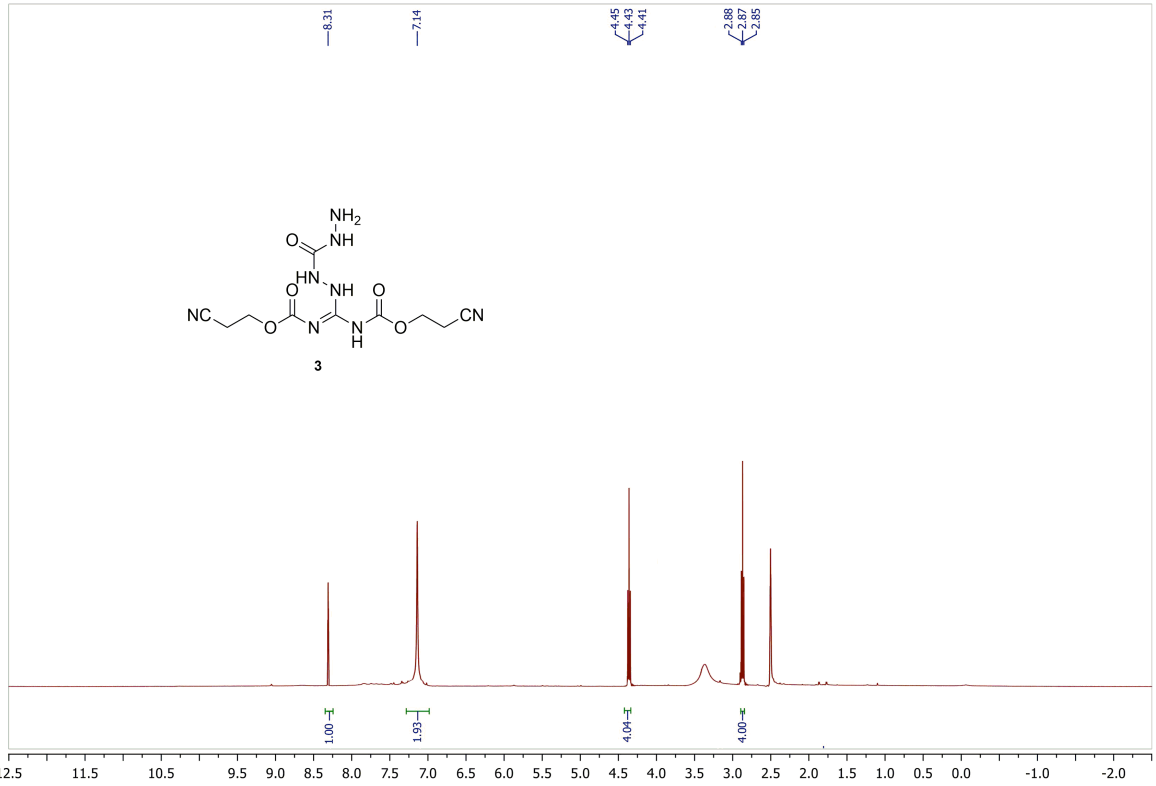


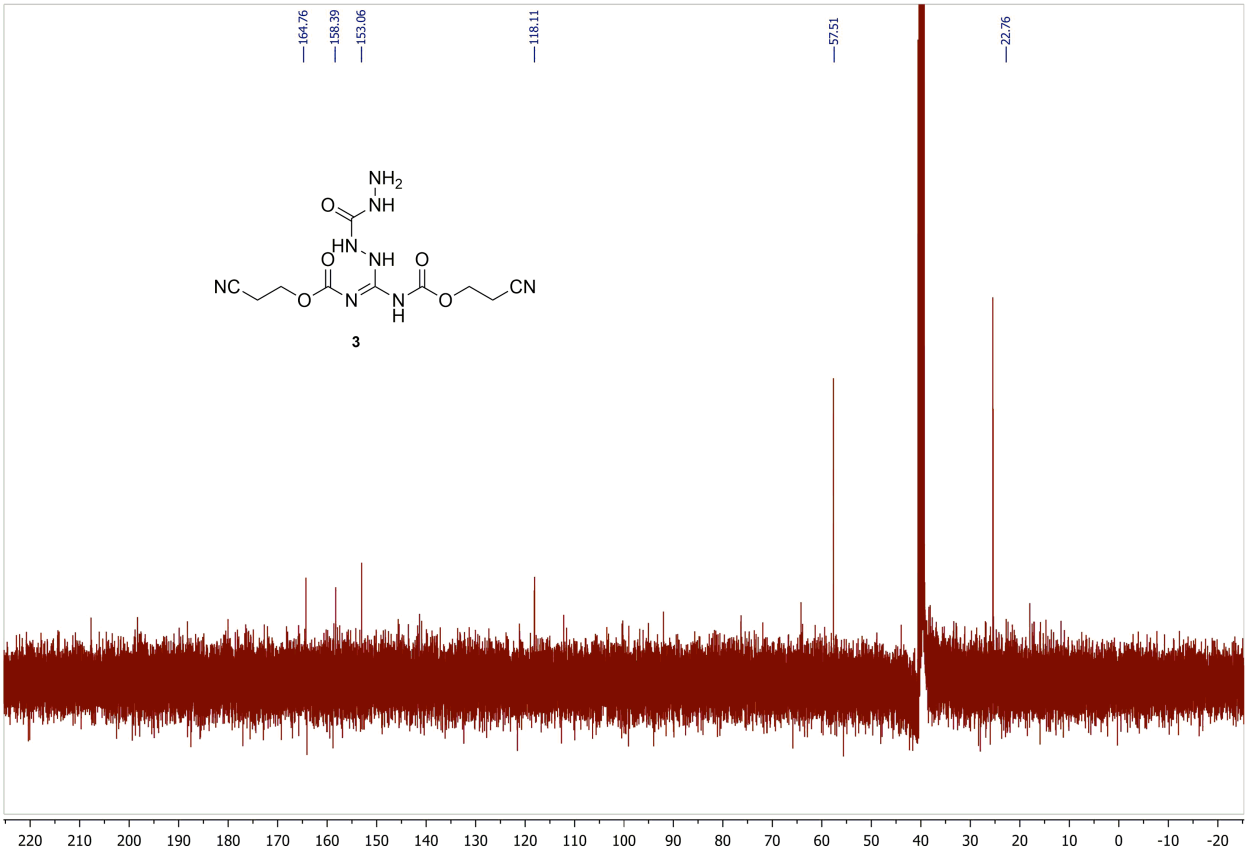
^{1}H and ^{13}C NMR for compound **2**



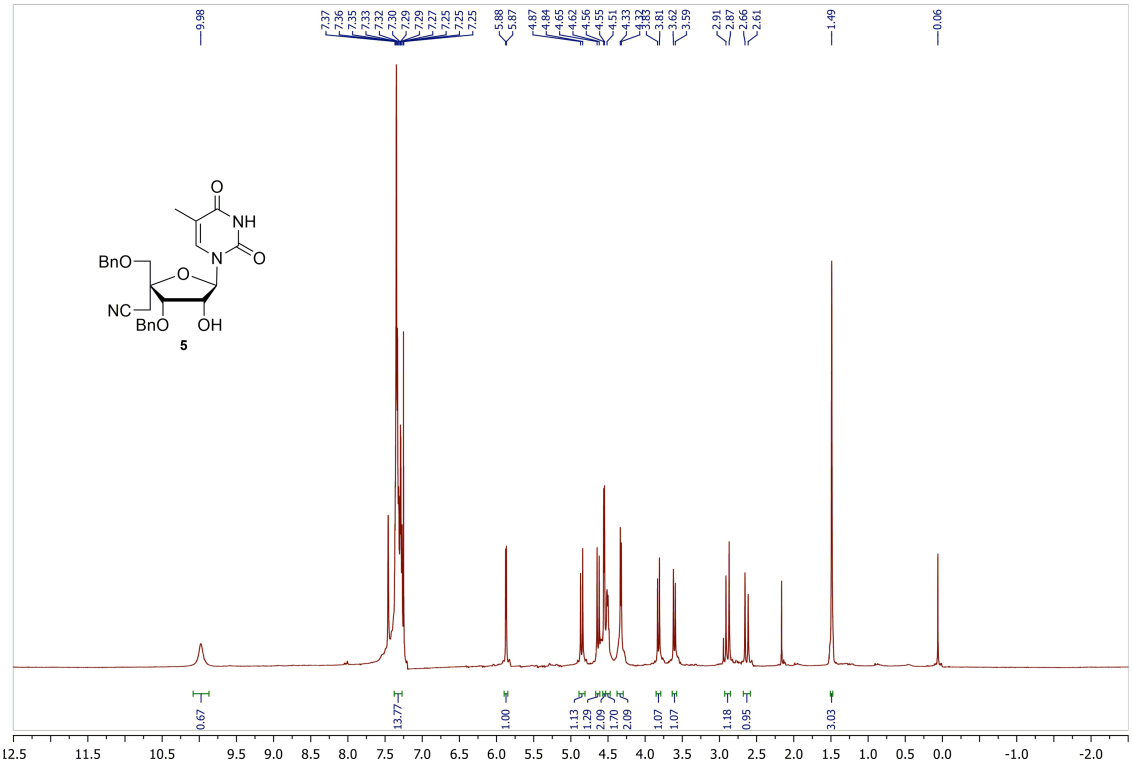


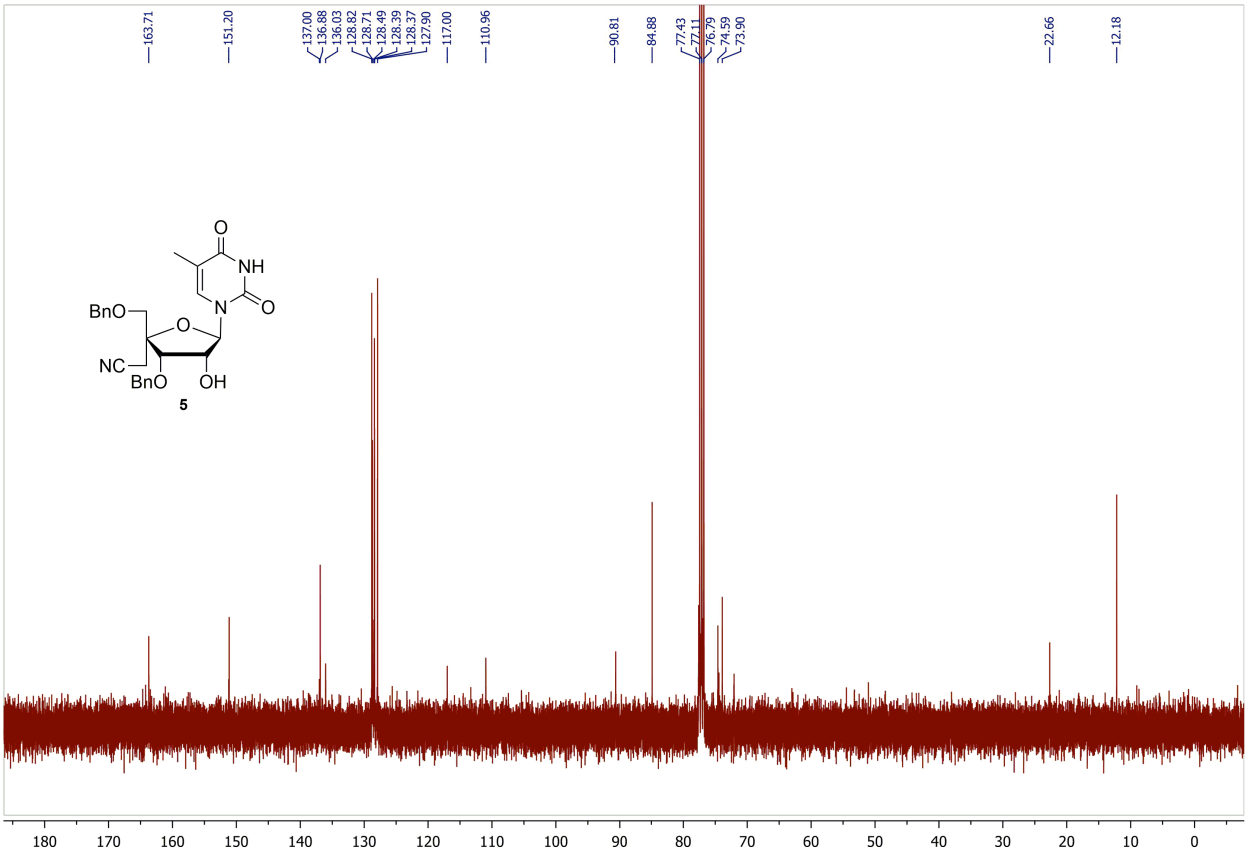
^{1}H and ^{13}C NMR for compound **3**



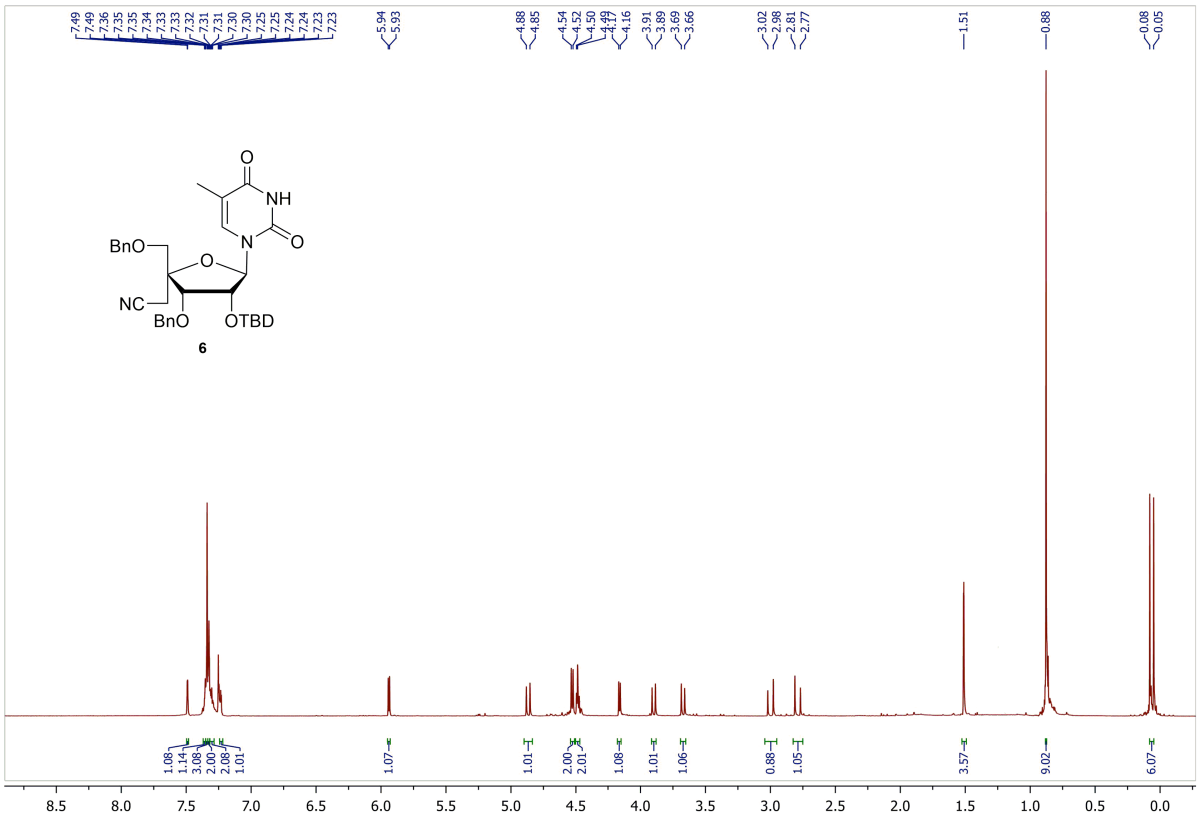


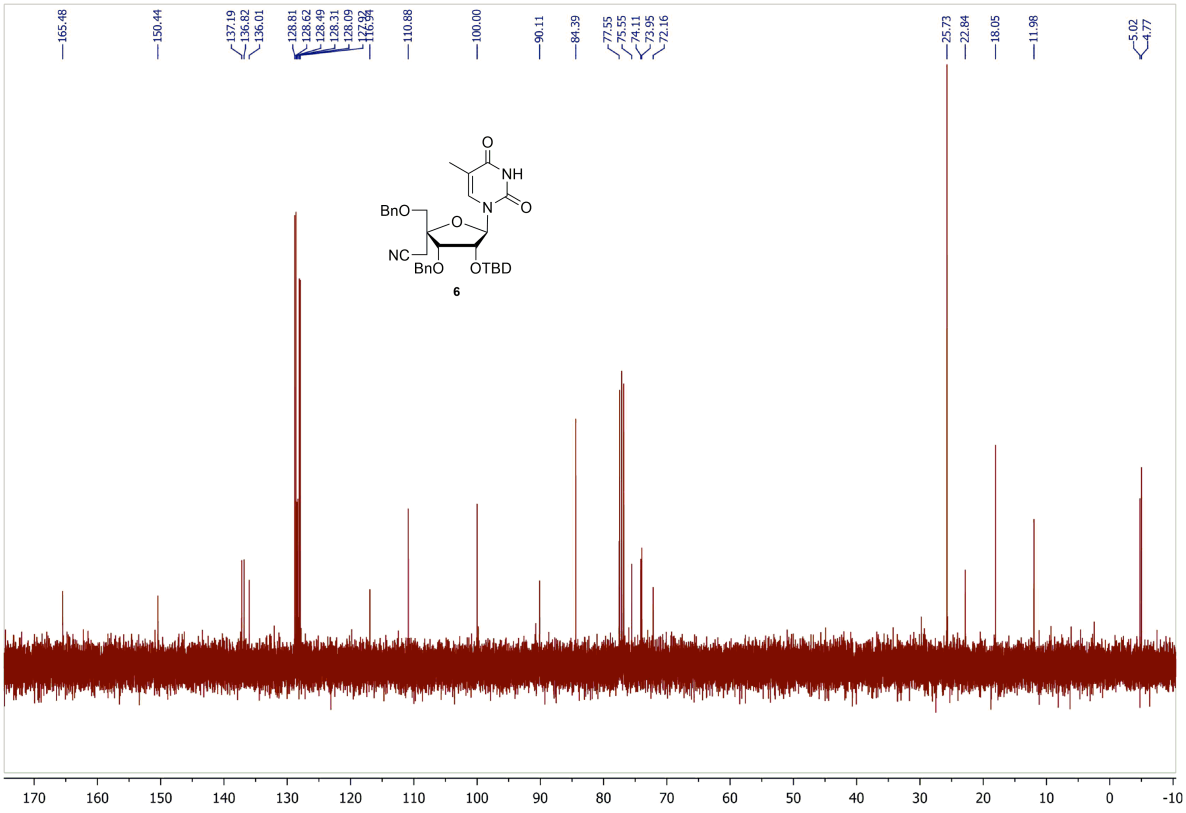
^{1}H and ^{13}C NMR for compound **5**



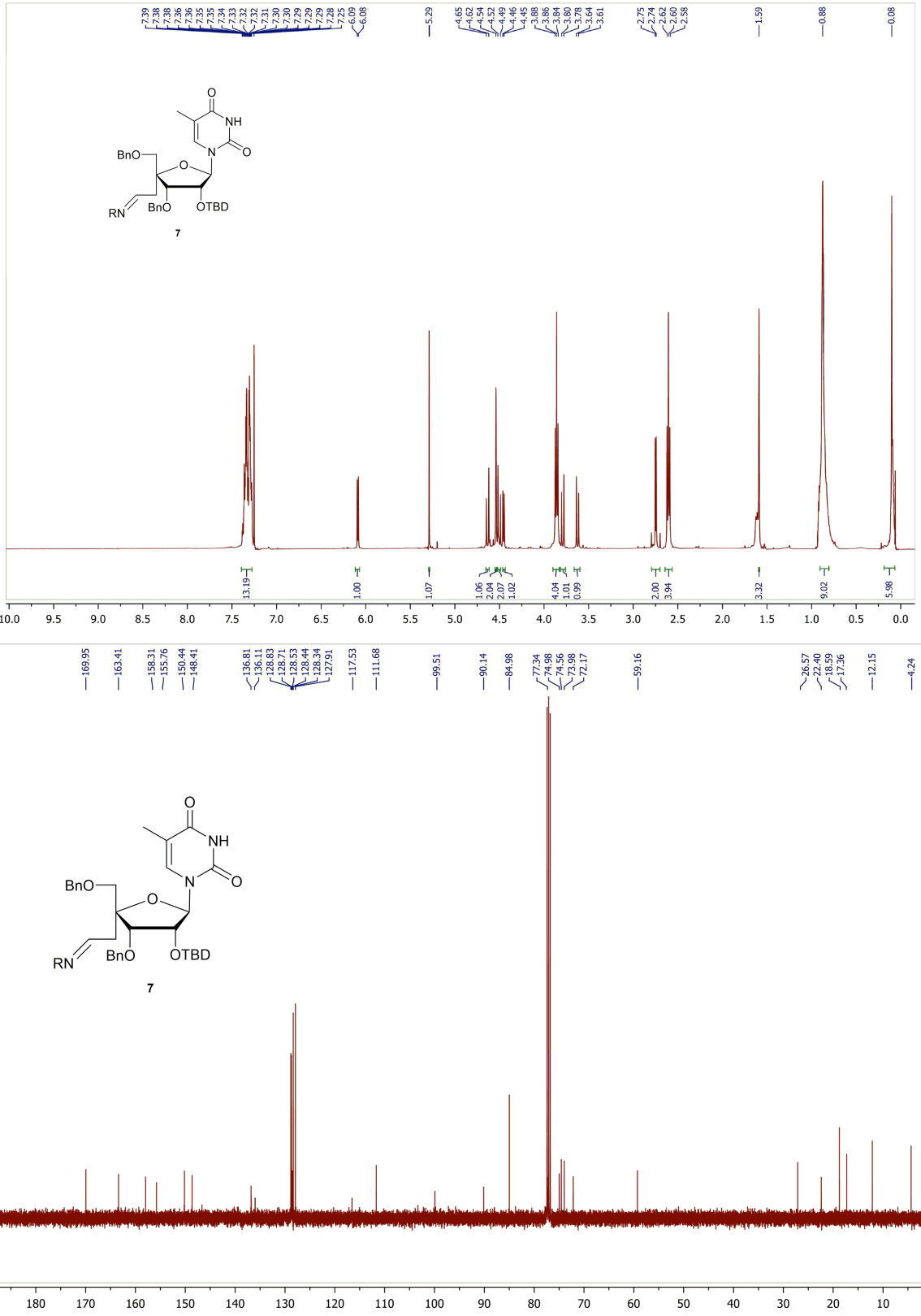


^{1}H and ^{13}C NMR for compound **6**

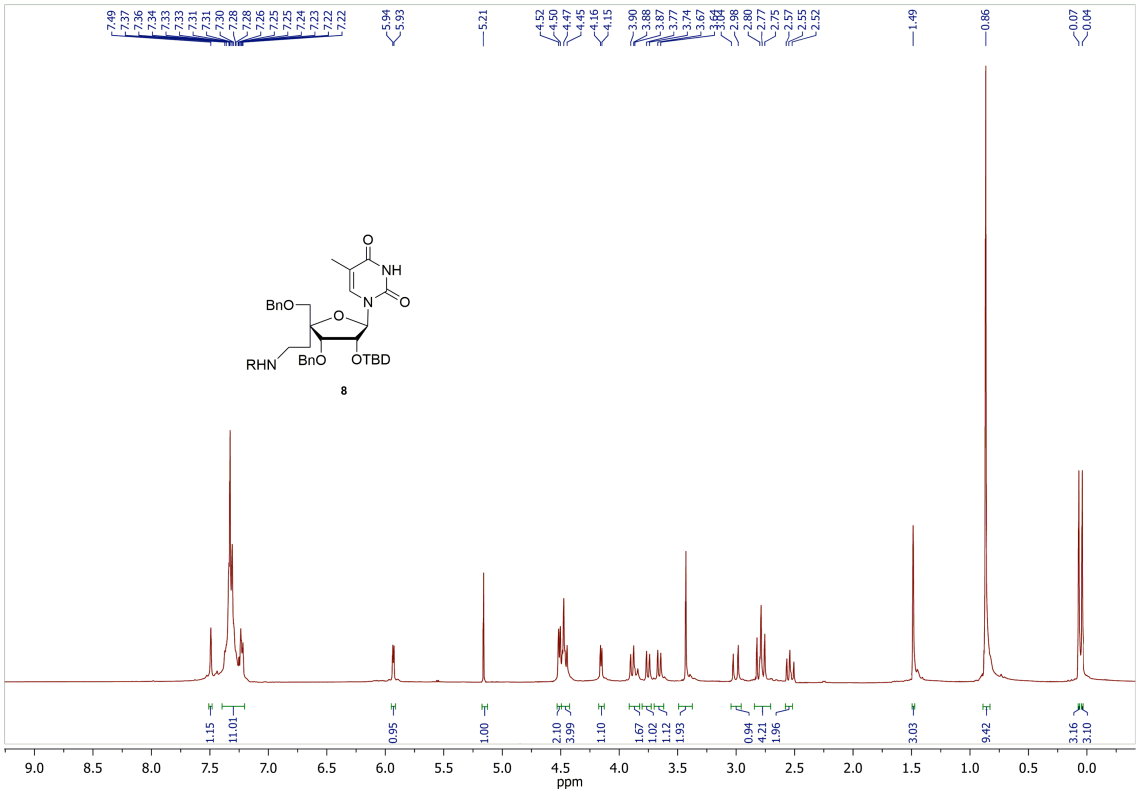


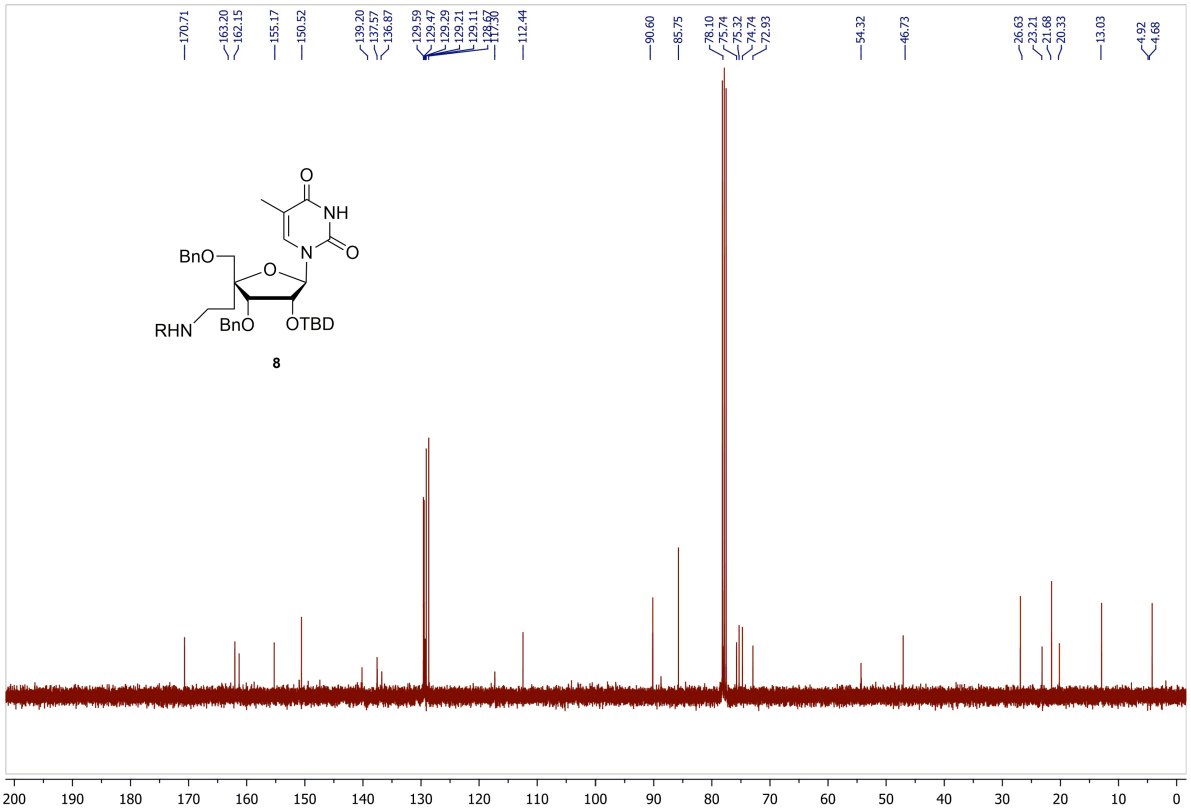


^{1}H and ^{13}C NMR for compound **7**

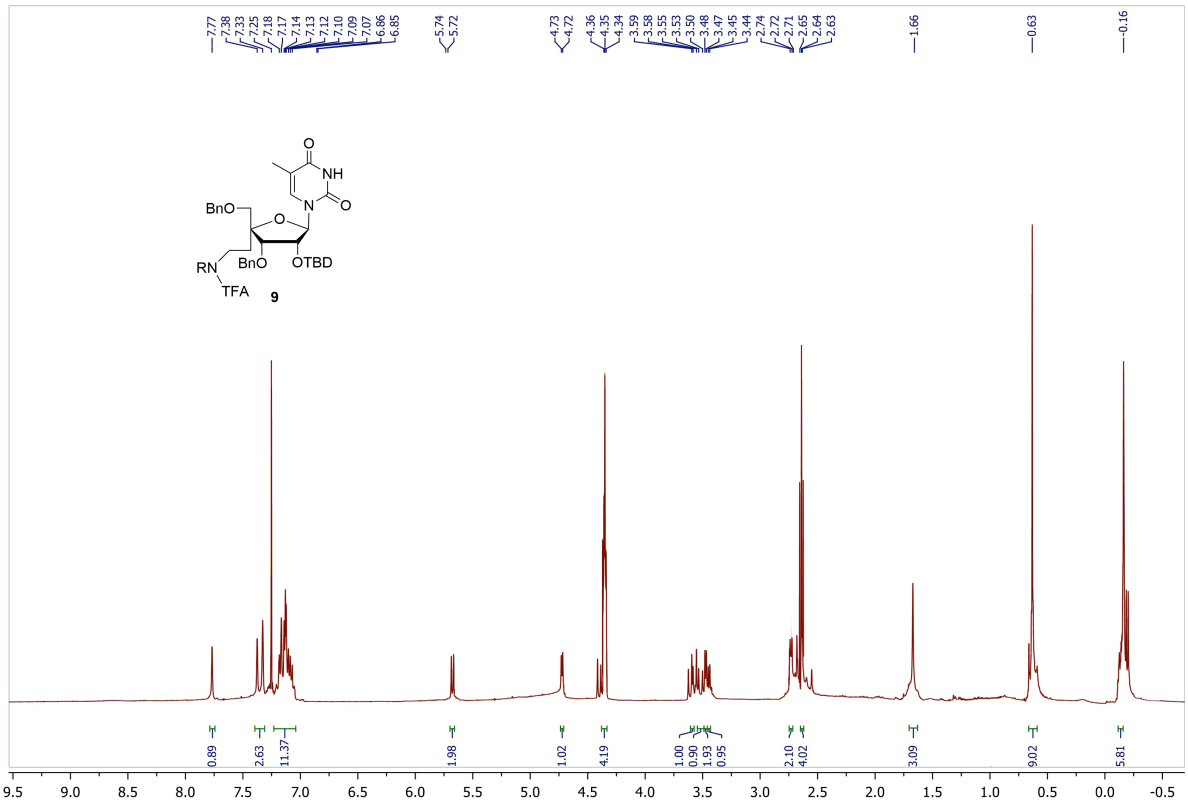


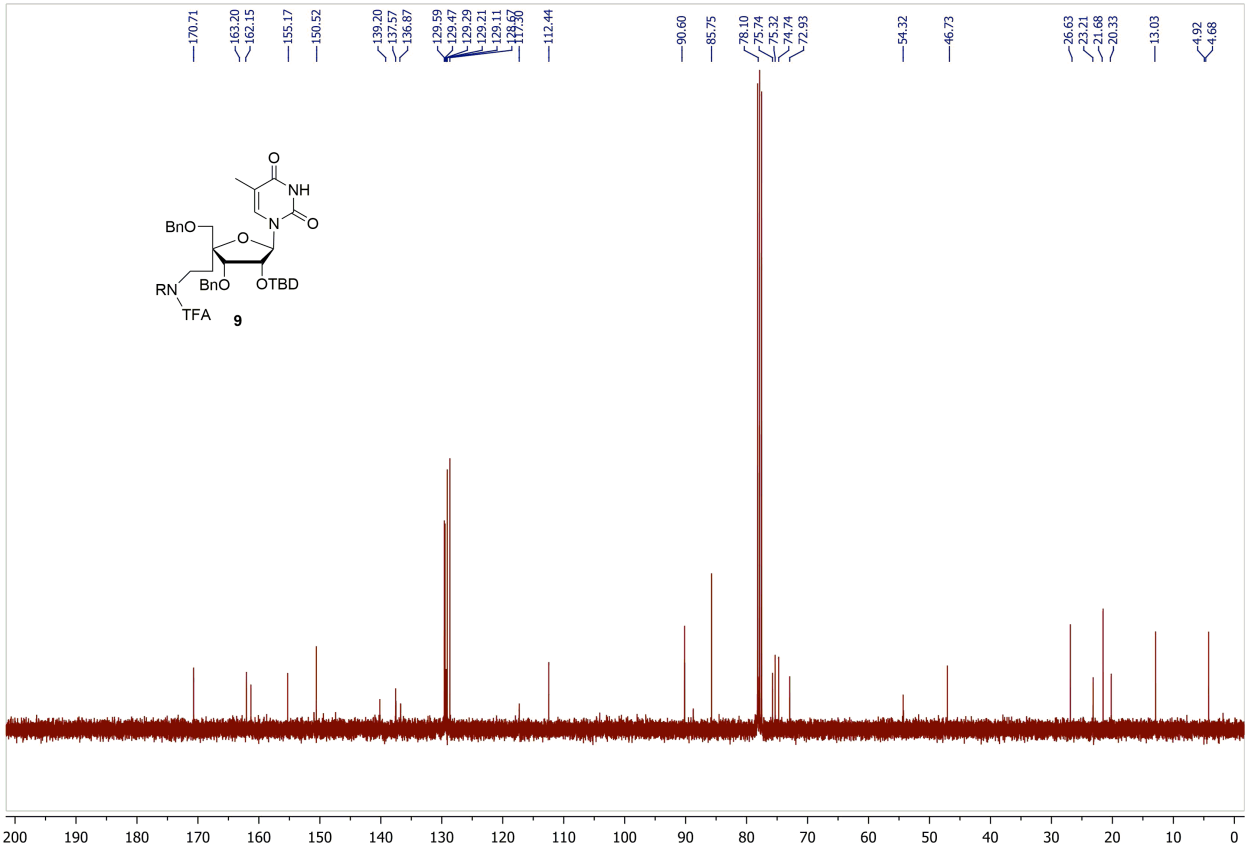
^{1}H and ^{13}C NMR for compound **8**

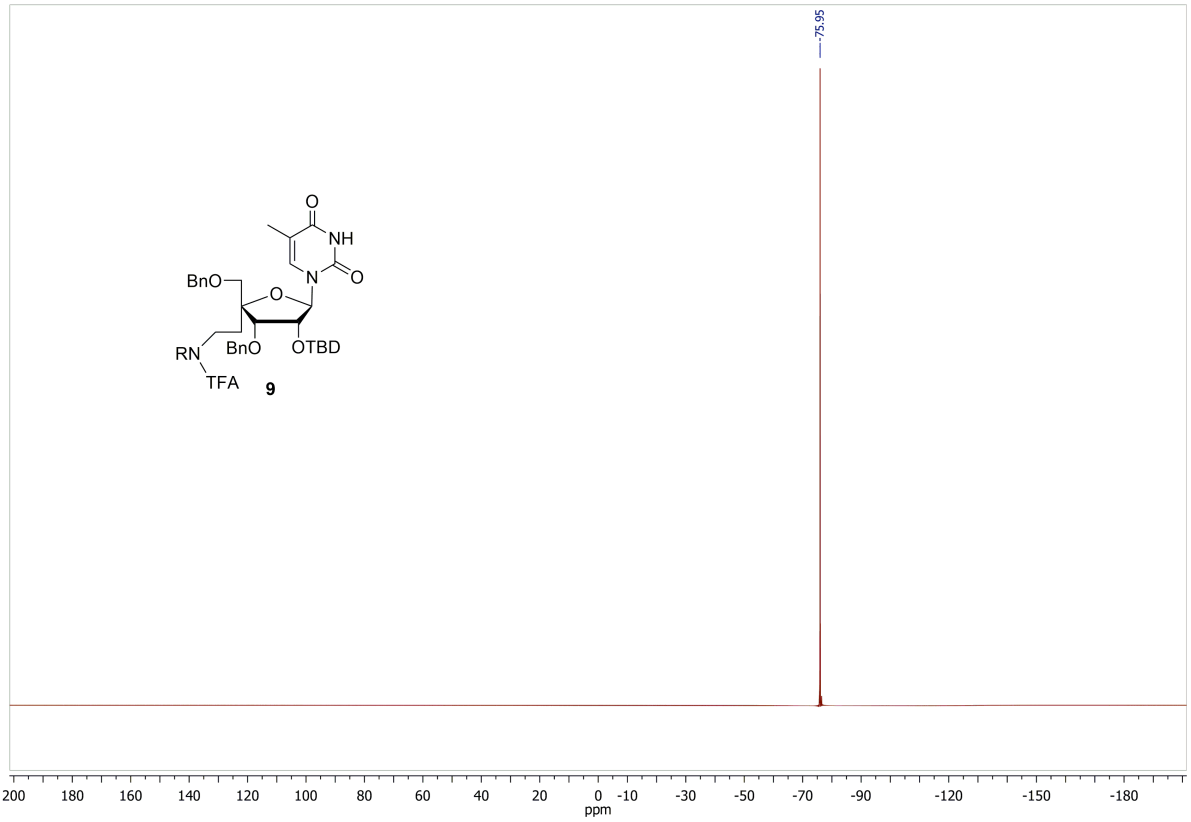

7.49
7.37
7.36
7.34
7.33
7.33
7.31
7.31
7.30
7.28
7.28
7.26
7.25
7.25
7.24
7.23
7.22
7.22
5.94
5.93
5.21
4.52
4.50
4.47
4.45
4.16
4.15
3.90
3.88
3.87
3.77
3.74
3.67
3.64
3.04
2.98
2.90
2.77
2.75
2.57
2.55
2.52
1.49
0.86
0.07
0.04
O
NH
N
O
BnO
O
RHN
BnO
OTBD
8
1.15
11.01
0.95
1.00
2.10
3.99
1.10
1.67
1.02
1.12
1.93
0.94
4.21
1.96
3.03
9.42
3.16
3.10
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
ppm


170.71
163.20
162.15
155.17
150.52
139.20
137.57
136.87
129.59
129.47
129.29
129.21
129.11
128.67
117.30
112.44
90.60
85.75
78.10
75.74
75.32
74.74
72.93
54.32
46.73
26.63
23.21
21.68
20.33
13.03
4.92
4.68
O
NH
N
O
BnO
O
RHN
BnO
OTBD
8
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0

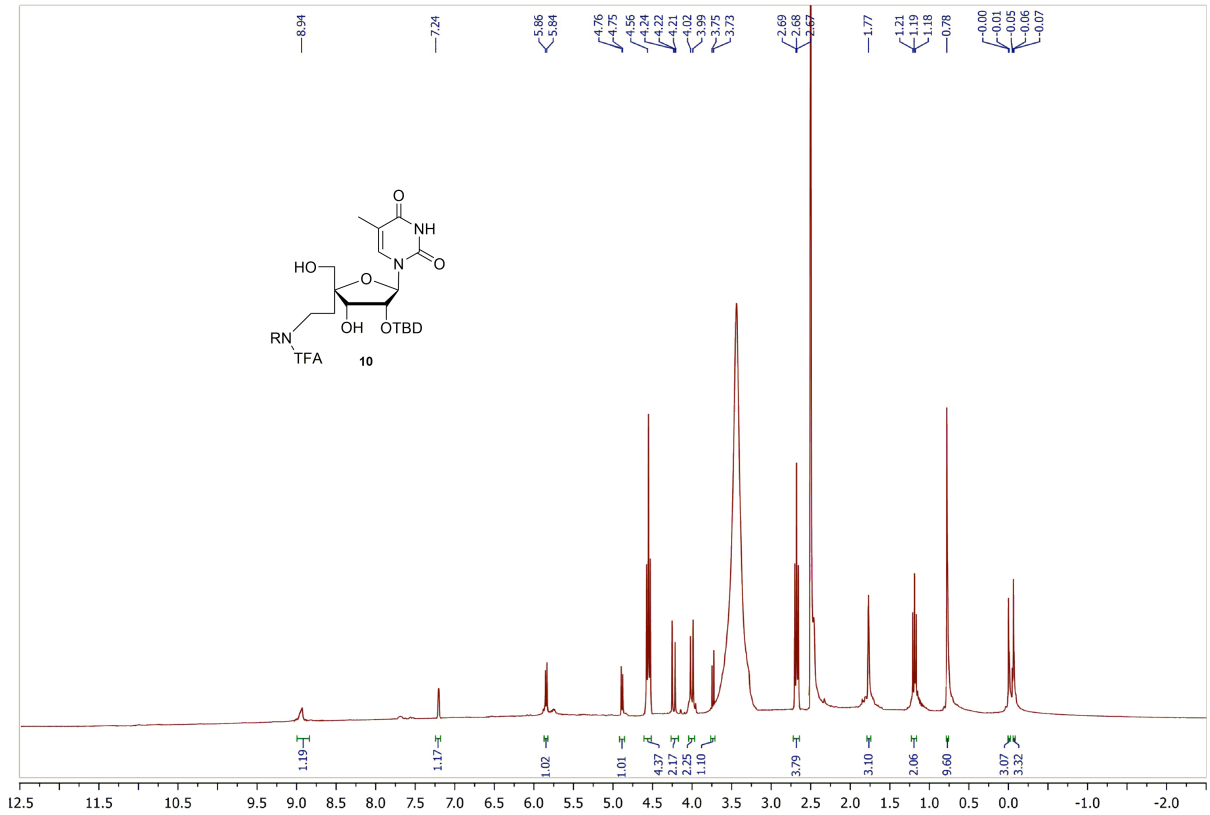
^{1}H, ^{13}C and ^{19}F NMR for compound **9**

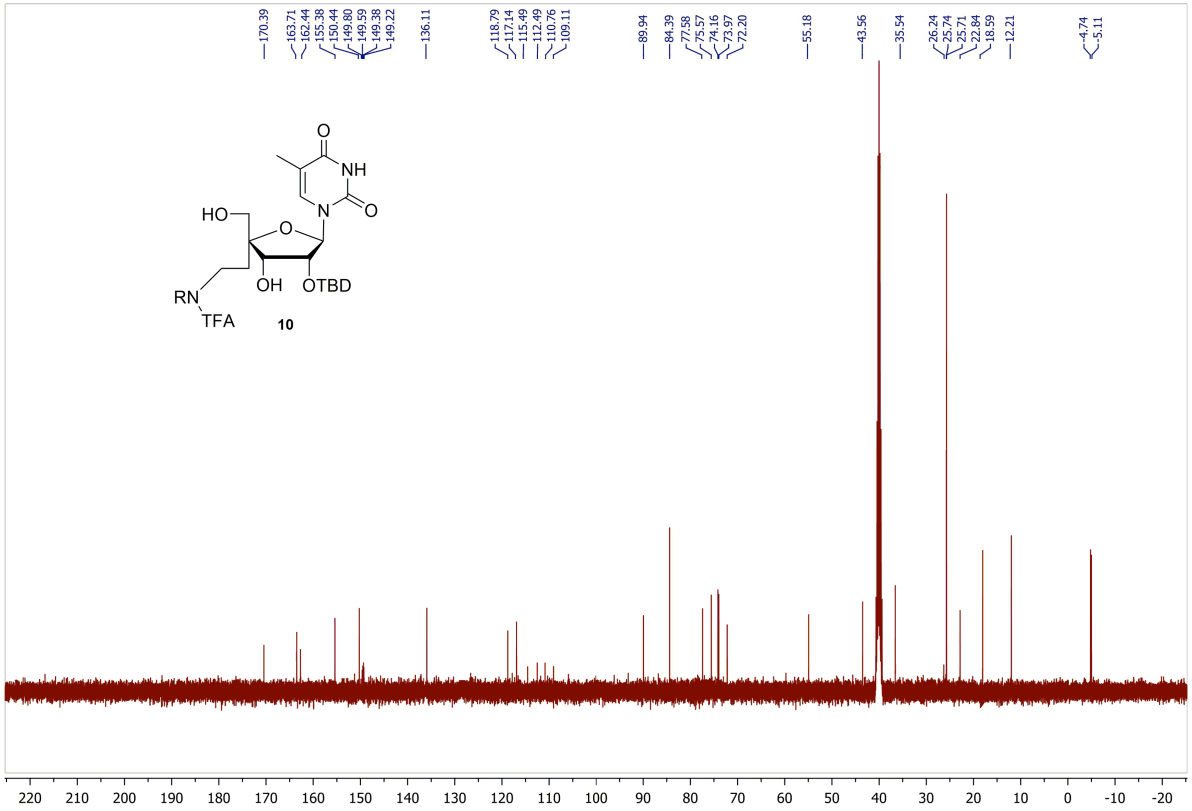


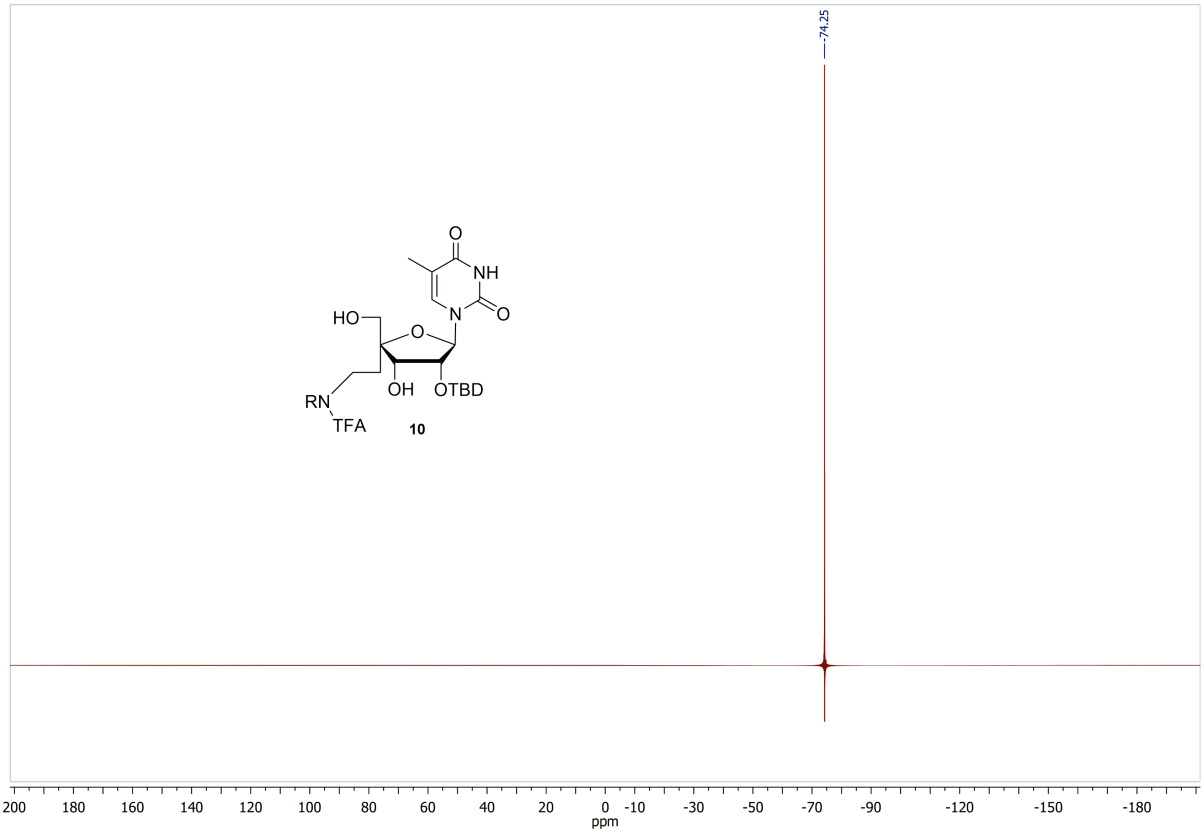



-75.95
O
NH
N
O
BnO
O
RN
BnO
OTBD
TFA
9
200
180
160
140
120
100
80
60
40
20
0
-10
-30
-50
-70
-90
-120
-150
-180
ppm

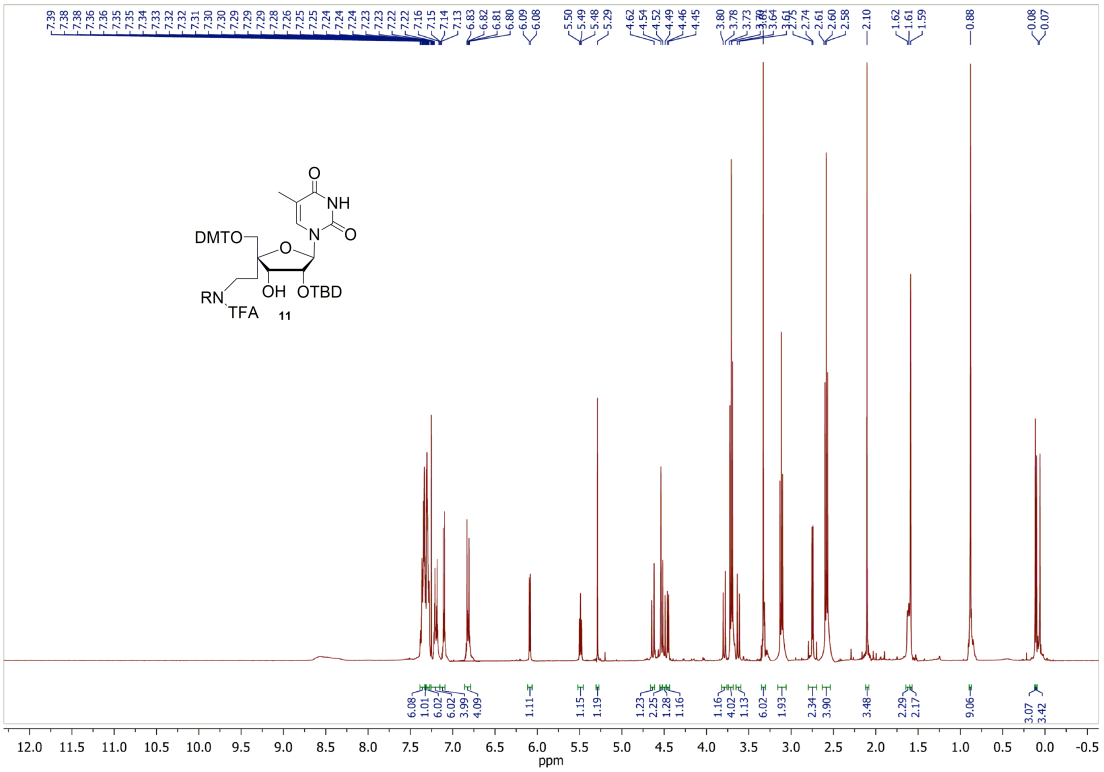
^{1}H, ^{13}C and ^{19}F NMR for compound **10**

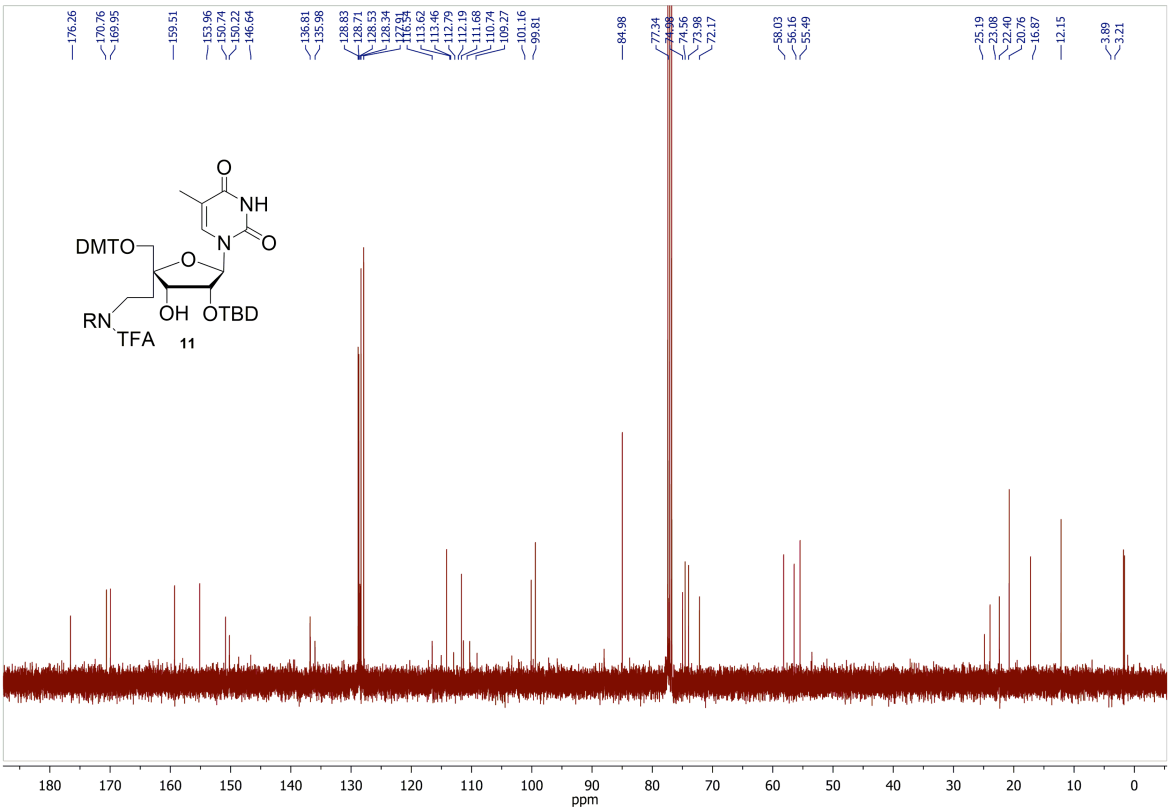


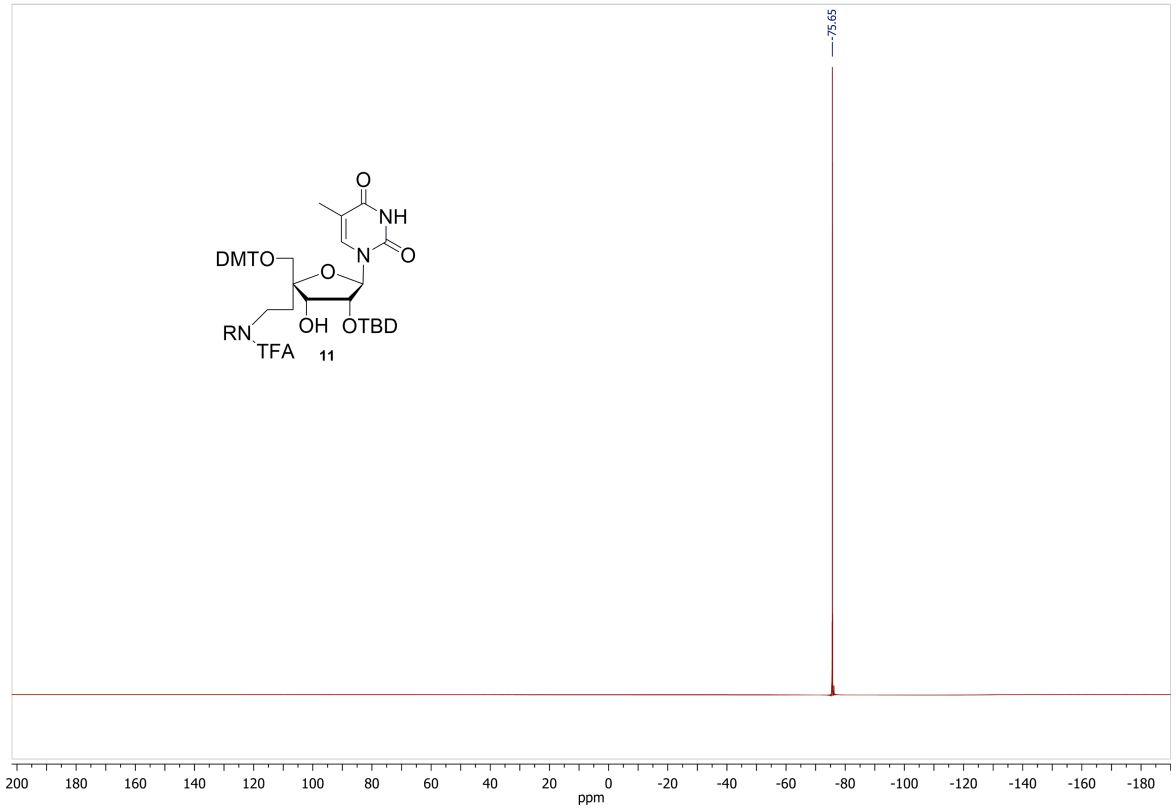



-74.25
O
NH
N
O
HO
O
OH
OTBD
RN
TFA
10
200
180
160
140
120
100
80
60
40
20
0
-10
-30
-50
-70
-90
-120
-150
-180
ppm

^{1}H, ^{13}C and ^{19}F NMR for compound **11**






-75.65
O
NH
N
O
DMTO
O
RN
TFA
OH
OTBD
11
200
180
160
140
120
100
80
60
40
20
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
ppm

^{31}P and ^{19}F NMR for compound **12**