Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry

Site-selected incorporation of 5-carboxymethylaminomethyl(-2-thio)uridine into RNA sequences by phosphoramidite chemistry

Grazyna Leszczynska*, Jakub Pięta, Karolina Wozniak and Andrzej Malkiewicz

Institute of Organic Chemistry, Lodz University of Technology, 90-924 Lodz, Poland.
Fax: (+48) 42 636 55 30
E-mail: grazyna.leszczynska@p.lodz.pl

Supporting Information

1. General.

NMR spectra were recorded on a Bruker Avance DPX 250 spectrometer at 250.0 (^{1}H), 62.9 (^{13}C), and 101.3 (^{31}P) or on a Bruker Avance II Plus 700 spectrometer at 700.0 (^{1}H) and 176.0 (^{13}C). Chemical shifts are reported in ppm relative to TMS (internal standard) for ^{1}H and ^{13}C, and 85% phosphoric acid (external standard) for ^{31}P. Chemical shifts are described as s (singlet), d (doublet), dd (doublet of doublets), t (triplet), q (quartet), m (multiplet), and bs (broad singlet). Coupling constants (*J*) are reported in Hertz. High-resolution mass spectra were obtained from a Finnigan MAT 95 spectrometer (FAB ionization) and MALDI SYNAPT G2-S HDMS (ESI ionization). Thin layer chromatography was done on Merck 60F_{254} coated plates, and Merck silica gel 60 (230–400 mesh) was used for column chromatography. HPLC was performed with a Waters chromatograph interfaced with a 996 spectral diode array detector. MALDI-TOF spectra were recorded on Applied Biosystems Voyager-Elite mass spectrometer.

2. Procedures for preparation of compounds 11a/11b, 12a/12b, 13a/13b

5'-*O*-(4,4'-dimethoxytrityl)-*N*-[(1-*β*-D-ribofuranosyl-1*H*-(-2-thio)pyrimidin-5-yl)methyl]-*N*-trifluoroacetylglycine 2-(trimethylsilyl)ethyl ester (11a/11b)

Nucleoside **10a**/**10b**[1] (1 mmol, 1.0 equiv) was dissolved in 7 mL of anhydrous pyridine. 4,4'-Dimethoxytrityl chloride (0.5 g, 1.5 mmol, 1.5 equiv) was added, and the mixture was stirred at rt for 24 h. The reaction mixture was diluted with 40 mL of $CHCl_3$ and washed with water (2 × 15 mL). The organic layer was dried over $MgSO_4$ and the solvent was removed under reduced pressure. The resulting foam was purified by column chromatography affording products **11a**/**11b** as rotamers about the –NC(O)CF_3 amide bond (two chemical shifts were observed for some of the ^{1}H and ^{13}C NMR resonances; the secondary shifts in ^{13}C NMR spectra are given in parentheses).

Compound **11a** was isolated by flash chromatography using 3% MeOH in $CHCl_3$ as eluent to obtain a white foam in 82% yield. TLC R_f = 0.43 ($CHCl_3$/MeOH 9:1, v/v); ^{1}H NMR (700 MHz; C_6D_6): δ -0.09 (s, 0.9H), 0.00 (s, 8.1H), 0.85–0.87 (m, 2H), 2.04 (bs, 1H), 3.40 (s, 0.7H), 3.42 (s, 5.3H), 3.63–3.68 (m, 2H), 3.91–4.67 (m, 11H), 5.38 (bs, 0.8H), 6.02 (d, 0.9H, *J* = 2.10 Hz), 6.11 (d, 0.1H, *J* = 3.50 Hz), 6.80–7.70 (m, 13H), 7.78 (s, 0.1H), 8.17 (s, 0.9H), 10.59 (bs, 0.8H); ^{13}C NMR (176 MHz, C_6D_6): δ -1.71 (-1.68), 17.39 (17.34), 47.37, 51.27, 54.87, 63.43, 64.22 (64.16), 70.68, 75.60, 84.22, 87.35 (87.30), 91.26, 109.28, 113.79 (113.71), 116.78 (q, *J* = 288.46 Hz), 127.24, 128.31 (128.27), 128.87 (128.75), 130.78 (130.62), 136.20, 136.37, 142.44, 145.47, 151.34 (151.45), 157.32–158.14 (m), 159.25 (159.34), 164.66, 168.84; HRMS calcd for $C_{40}H_{46}N_3O_{11}F_3NaSi$ $[M+Na]^+$ 852.2751, found 852.2742.

Compound **11b** was isolated by flash chromatography using $CHCl_3$ as eluent to obtain a yellow foam in 84% yield. TLC R_f = 0.66 ($CHCl_3$/MeOH 95:5, v/v); ^{1}H NMR (700 MHz; $(CD_3)_2CO$): δ 0.07 (s, 1.8H), 0.09 (s, 7.2H), 0.97–0.99 (m, 0.4H), 1.04–1.06 (m, 1.6H), 3.45–3.71 (m, 2H), 3.82 (s, 4.8H), 3.83 (s, 1.2H), 3.86 (d, 1H, *J* = 14.70 Hz), 3.98 (d, 1H, *J* = 14.70 Hz), 4.07–4.55 (m, 7H), 4.83 (bs, 0.9H), 6.57 (d, 0.8H, *J* = 2.10 Hz), 6.64 (d, 0.2H, *J* = 2.10 Hz), 7.26–7.60 (m, 13H), 7.83 (s, 0.2H), 8.01 (s, 0.8H), 11.47 (bs, 0.8H); ^{13}C NMR (176 MHz, $(CD_3)_2CO$): δ -0.49 (-0.47), 18.87 (18.84), 48.41, 52.28, 56.48, 65.33 (65.19), 65.54, 71.56 (71.67), 76.95 (76.68), 85.28 (85.09), 88.29 (88.33), 96.35 (96.11), 115.02 (114.99), 118.08 (q, *J* = 287.76 Hz), 128.64 (128.80), 129.69 (129.74), 130.19 (130.11), 132.15 (132.06), 132.16 (132.01), 142.82, 147.04 (146.82), 158.50 (q, *J* = 32.38 Hz, 160.68 (160.78), 162.23, 170.19 (169.60), 177.51; HRMS calcd for $C_{40}H_{46}N_3O_{10}F_3NaSiS$ $[M+Na]^+$ 868.2523, found 868.2526.

5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)-*N*-[(1-*β*-D-ribofuranosyl-1*H*-(-2-thio)pyrimidin-5-yl)methyl]-*N*-trifluoroacetylglycine 2-(trimethylsilyl)ethyl ester (12a/12b)

The 5'-DMT nucleoside **11a**/**11b** (0.72 mmol, 1.0 equiv) was dissolved in 6 mL of anhydrous pyridine, and then imidazole (0.15 g, 2.2 mmol, 3.0 equiv) and *t*-butyldimethylsilyl chloride (0.16 mg, 1.0 mmol, 1.4 equiv) were added. After being stirred for 4 h at rt, the reaction mixture was diluted with 20 mL of $CHCl_3$ and washed with water (2 × 6 mL). The organic layer was dried over $MgSO_4$ and the solvent removed under reduced pressure. An equimolar mixture of 2'- and 3'-TBS isomers was separated by column chromatography. The 2D COSY NMR experiment was used to confirm the identity of the 2' isomer from the correlation of H3' with 3'OH. To obtain additional quantities of the 2' TBS isomer **12a**/**12b**, the 3' isomer was isomerized to an equimolar mixture of 2' and 3' isomers by stirring in methanol with a trace of triethylamine.

Compound **12a** was purified on a silica gel column with 16% acetone in DCM as eluent to obtain a white foam in 70% yield. TLC R_f = 0.70 (DCM/acetone 9:1, v/v); ^{1}H NMR (700 MHz, $(CD_3)_2SO$): δ -0.01-0.03 (m, 9H), 0.02-0.04 (m, 6H), 0.82-0.84 (m, 9H), 0.86-0.92 (m, 2H), 3.20-3.27 (m, 2H), 3.70-3.73 (m, 6H), 3.84-4.41 (m, 9H), 5.06 (d, 0.02H, *J* = 6.3 Hz), 5.09 (d, 0.01H, *J* = 6.3 Hz), 5.71 (d, 0.34H, *J* = 3.5 Hz), 5.75 (d, 0.66H, *J* = 4.90 Hz), 6.86-6.88 (m, 4H), 7.18-7.42 (m, 9H), 7.54 (s, 0.34H), 7.71 (s, 0.66H); ^{13}C NMR (176 MHz, C_6D_6): δ -5.41, -4.75, -1.92, 17.17, 17.95, 25.64, 47.11, 51.08, 54.55, 63.70 (63.63), 63.80, 70.44, 75.95, 83.93, 87.08, 89.84, 108.58, 113.24, 113.54, 113.57, 116.59 (q, *J* = 288.50 Hz), 127.00, 128.00 (128.05), 128.66, 129.46, 130.54, 130.62, 136.14, 141.77, 145.36, 149.91, 157.304 (q, *J* = 35.55 Hz), 159.09 (159.05), 164.00, 168.44; HRMS calcd for $C_{46}H_{60}N_3O_{11}F_3NaSi_2$ $[M+Na]^+$ 966.3616, found 966.3617.

Compound **12b** was purified on a silica gel column with $CHCl_3$ as eluent to obtain a light yellow foam in 65 % yield. TLC R_f = 0.62 (DCM/acetone 98:2, v/v); ^{1}H NMR (700 MHz; $(CD_3)_2SO$): δ -0.01-0.00 (m, 9H), 0.04-0.07 (m, 6H), 0.85-0.86 (m, 9H), 0.88-0.90 (m, 2H), 3.12-3.40 (m, 2H), 3.71-3.72 (m, 6H), 3.77-4.44 (m, 9H), 5.22 (d, 0.07H, *J* = 6.3 Hz), 5.25 (d, 0.24H, *J* = 5.6 Hz), 6.51 (d, 0.70H, *J* = 2.8 Hz), 6.54 (d, 0.30H, *J* = 2.10 Hz), 6.84-6.86 (m, 4H), 7.17-7.31 (m, 7H), 7.37-7.43 (m, 2H), 7.54 (s, 0.3H), 7.75 (s, 0.7H); ^{13}C NMR (176 MHz, $(CD_3)_2CO$): δ -3.31 (-3.25), -0.48 (-0.45), 18.87 (18.85), 19.89 (19.93), 27.36, 48.53, 52.42, 56.48, 65.55 (65.33), 65.79, 72.21 (72.09), 78.69 (78.47), 85.35, 88.39, 95.62 (95.95), 114.61 (114.52), 114.99 (115.01), 115.03 (115.06), 118.09 (q, *J* = 288.29 Hz), 128.41, 128.66, 129.28, 129.71 (129.74), 129.77, 130.18 (130.13), 131.04, 132.13, 132.18, 137.58, 137.78, 142.87, 147.02 (146.77), 150.36, 158.60 (q, *J* = 36.96 Hz), 160.69 (160.79), 162.08, 170.17 (169.58), 177.81 (178.08); HRMS calcd for $C_{46}H_{60}N_3O_{10}F_3NaSi_2S$ $[M+Na]^+$ 982.3388, found 982.3401.

5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)-*N*-[(1-*β*-D-ribofuranosyl-1*H*-(-2-thio)pyrimidin-5-yl)methyl]-*N*-trifluoroacetylglycine 2-(trimethylsilyl)ethyl ester 3'-(cyanoethyl *N*,*N*-diisopropylphosphoramidite) (13a/13b)

The 5'-DMT, 2'-TBS nucleoside **12a**/**12b** (0.19 mmol, 1.0 equiv) was dissolved in 1.4 mL anhydrous DCM under Ar atmosphere, then diisopropylethylamine (132 μL, 0.76 mmol, 4.0 equiv) and 2-cyanoethyl *N*,*N*-diisopropylphosphoramidic chloride (85 μL, 0.38 mmol, 2.0 equiv) were added. The reaction mixture was stirred at rt for 5 h (**12a**)/7 h (**12b**), diluted with 4 mL of DCM, and washed with 5% aq. $NaHCO_3$ (3 × 2 mL). The organic layer was dried over $MgSO_4$ and the solvent was removed under reduced pressure. The crude product **13a/13b** was purified by flash chromatography using petroleum ether:ethyl acetate (2:1 v/v). The material exists as a mixture of stereoisomers about phosphorus, wherein each is a rotamer

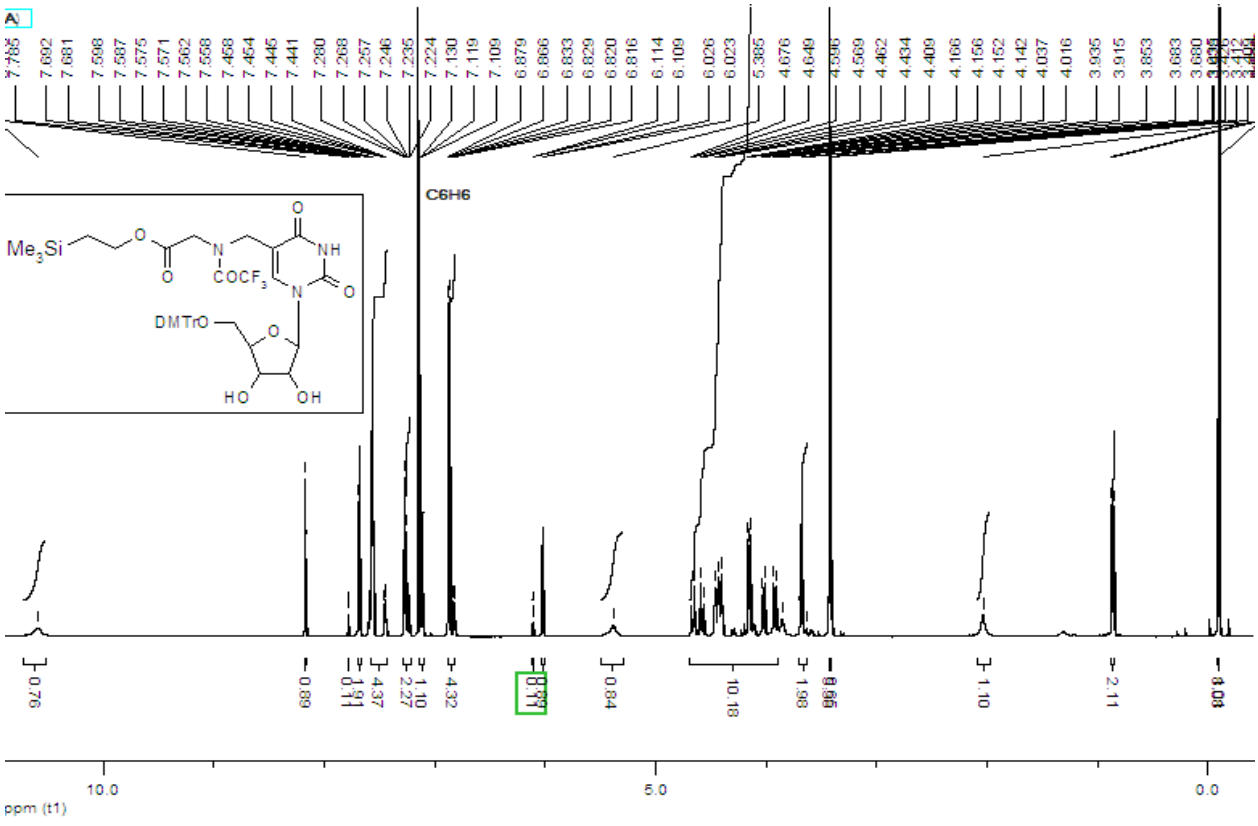
about the –NC(O)CF_3 amide bond and two or more chemical shifts are observed for some of the ^{1}H and ^{31}P NMR resonances.

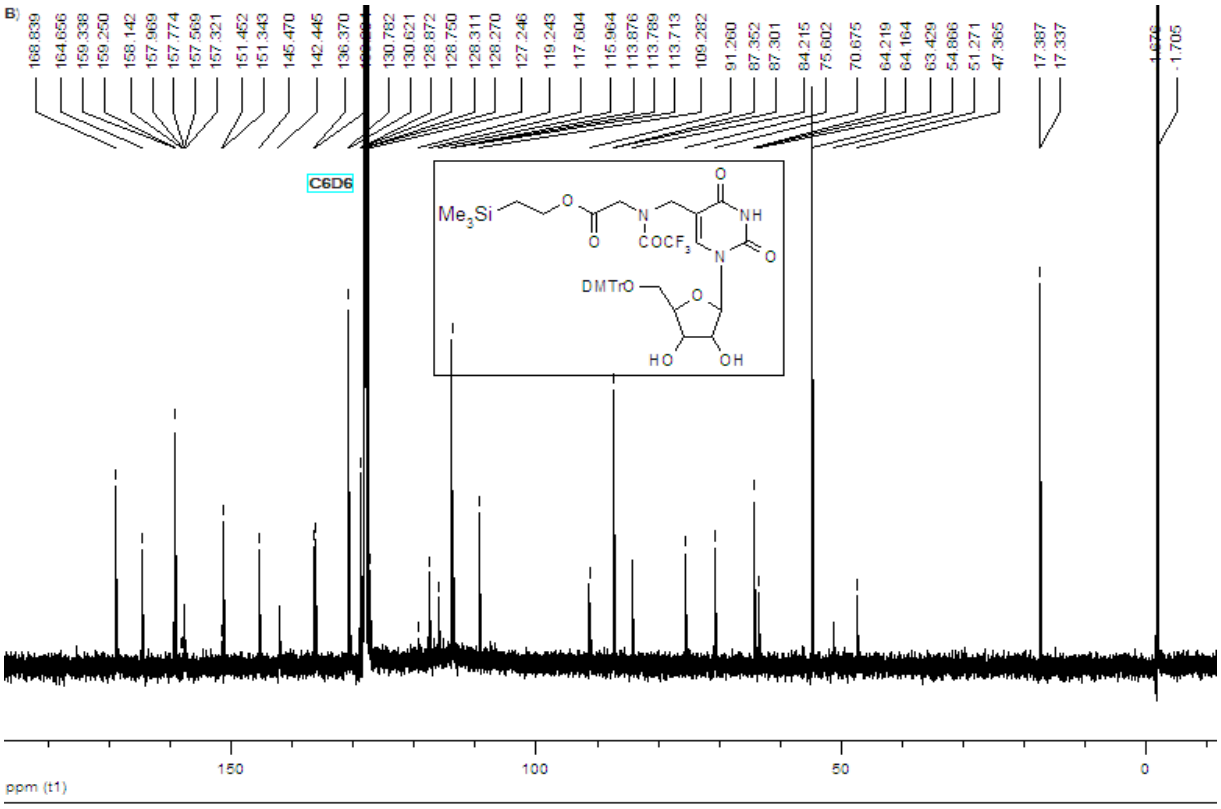
Compound **13a** was isolated as a white foam in 90% yield. TLC R_f = 0.60 (benzene, DCM, TEA 7:2:1 v/v/v); ^{1}H NMR (250 MHz; $(CD_3)_2CO$): δ -0.07–0.06 (m, 9H), 0.11–0.16 (m, 6H), 0.86–0.93 (m, 9H), 0.96–1.24 (m, 14H), 2.49–2.77 (m, 2H), 3.36–3.74 (m, 5H), 3.75–3.81 (m, 6H), 3.85–4.53 (m, 10H), 5.88–6.01 (m, 1H), 6.81–6.95 (m, 4H), 7.23–7.58 (m, 9H), 7.69–7.94 (m, 1H); ^{31}P NMR (101.25 Hz, $(CD_3)_2CO$): δ 149.89, 150.06, 150.80; HRMS calcd for $C_{55}H_{77}N_5O_{12}F_3NaSi_2P$ $[M+Na]^+$ 1166.4695, found 1166.4667.

Compound **13b** was isolated as a light yellow foam in 92% yield. TLC R_f = 0.70 (benzene, DCM, TEA 7:2:1 v/v/v); ^{1}H NMR (250 MHz; $(CD_3)_2CO$): δ -0.01-0.24 (m, 15H), 0.82-1.21 (m, 23H), 2.50-2.79 (m, 2H), 3.41-4.63 (m, 21H), 6.81-6.96 (m, 5H), 7.16-7.59 (m, 9H), 7.77-8.04 (m, 1H); ^{31}P NMR (101.25 Hz, $(CD_3)_2CO$): δ 150.28, 150.46, 150.57, 150.75; HRMS calcd for $C_{55}H_{77}N_5O_{11}F_3NaSi_2PS$ $[M+Na]^+$ 1182.4466, found 1182.4476.

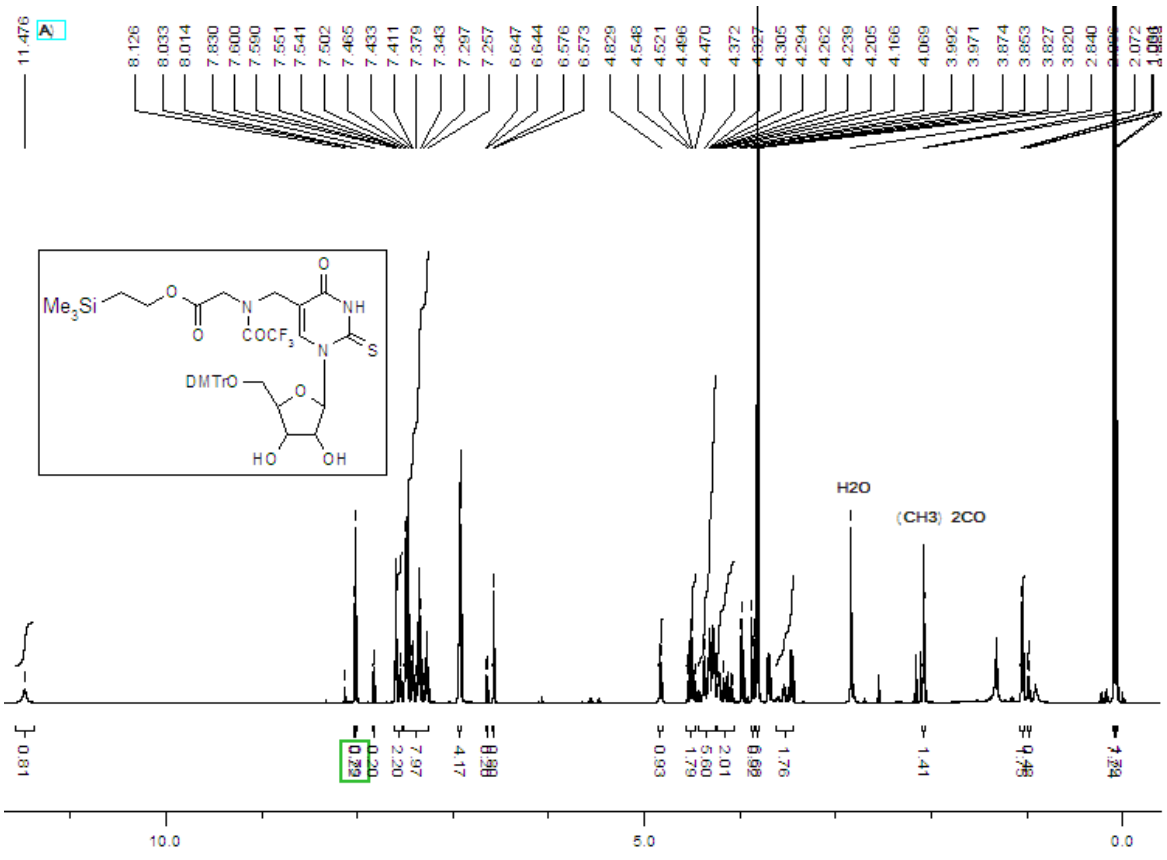
3. Spectra of compounds 11a/11b, 12a/12b, 13a/13b

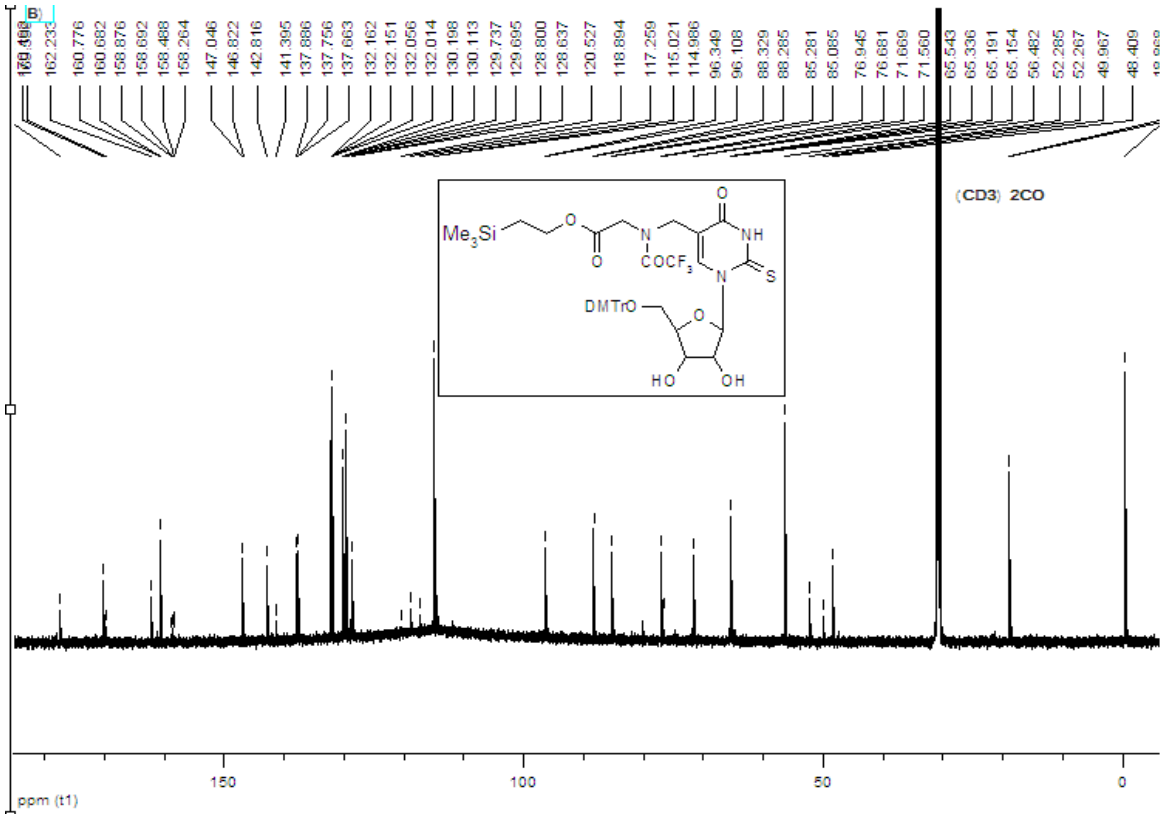
Supplementary Fig 1 A) ^{1}H NMR spectrum of **11a** (700 MHz, C_6D_6); **B)** ^{13}C NMR spectrum of **11a** (176 MHz, C_6D_6).



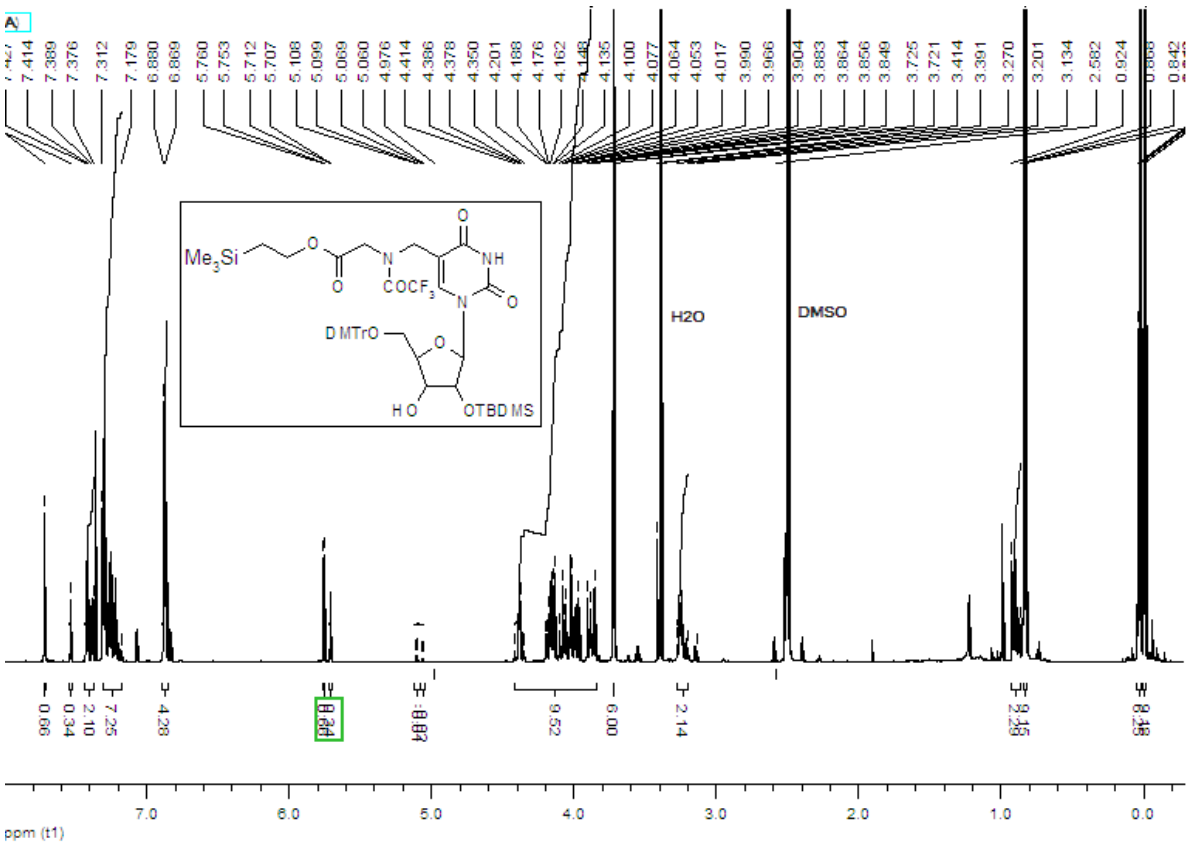


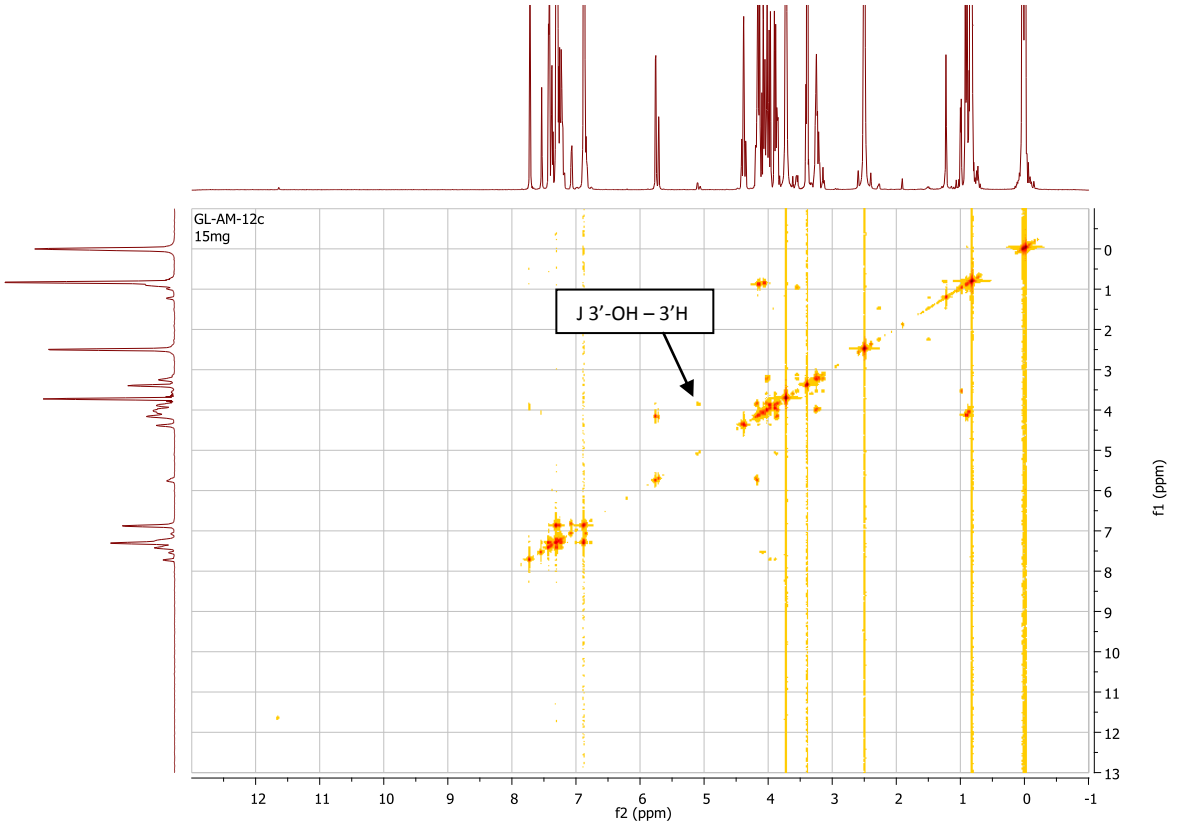
Supplementary Fig 2 A) ^{1}H NMR spectrum of **11b** (700 MHz, $(CD_3)_2CO$); **B)** ^{13}C NMR spectrum of **11b** (176 MHz, $(CD_3)_2CO$).

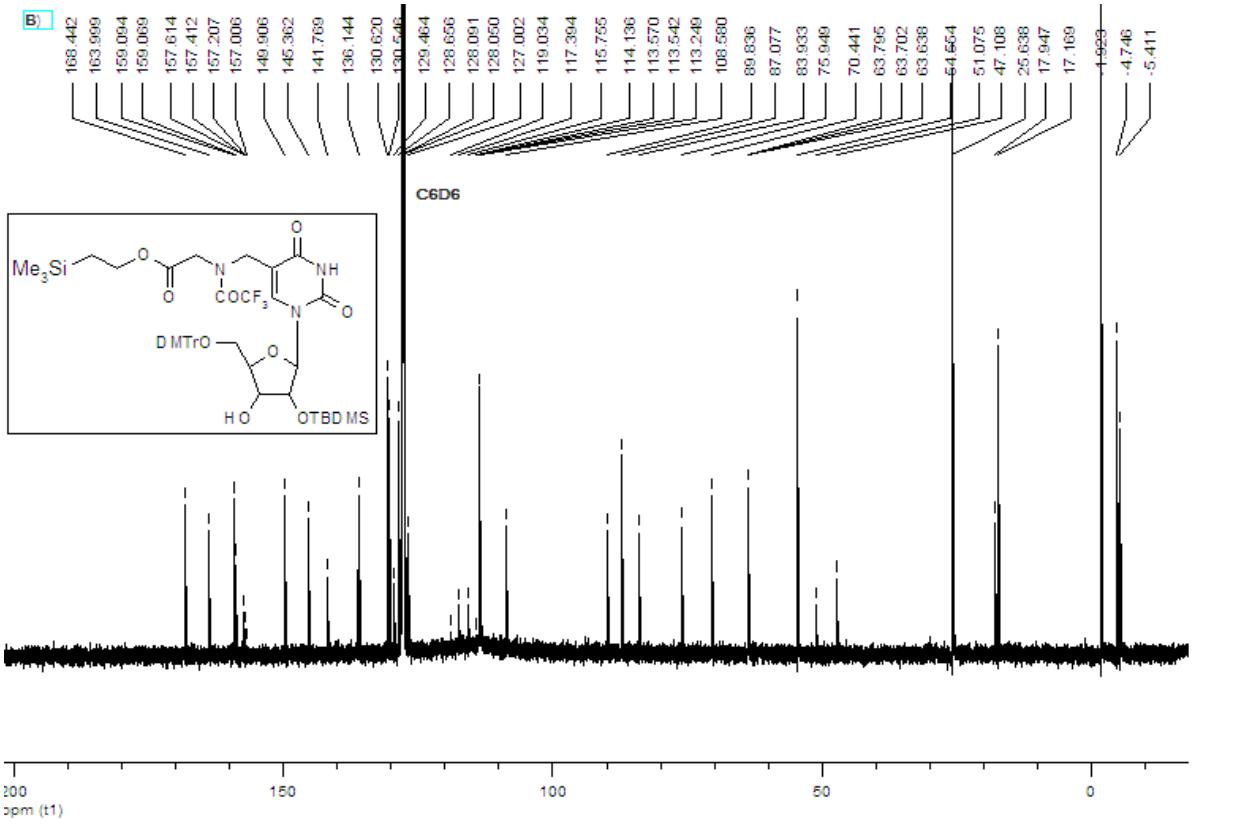




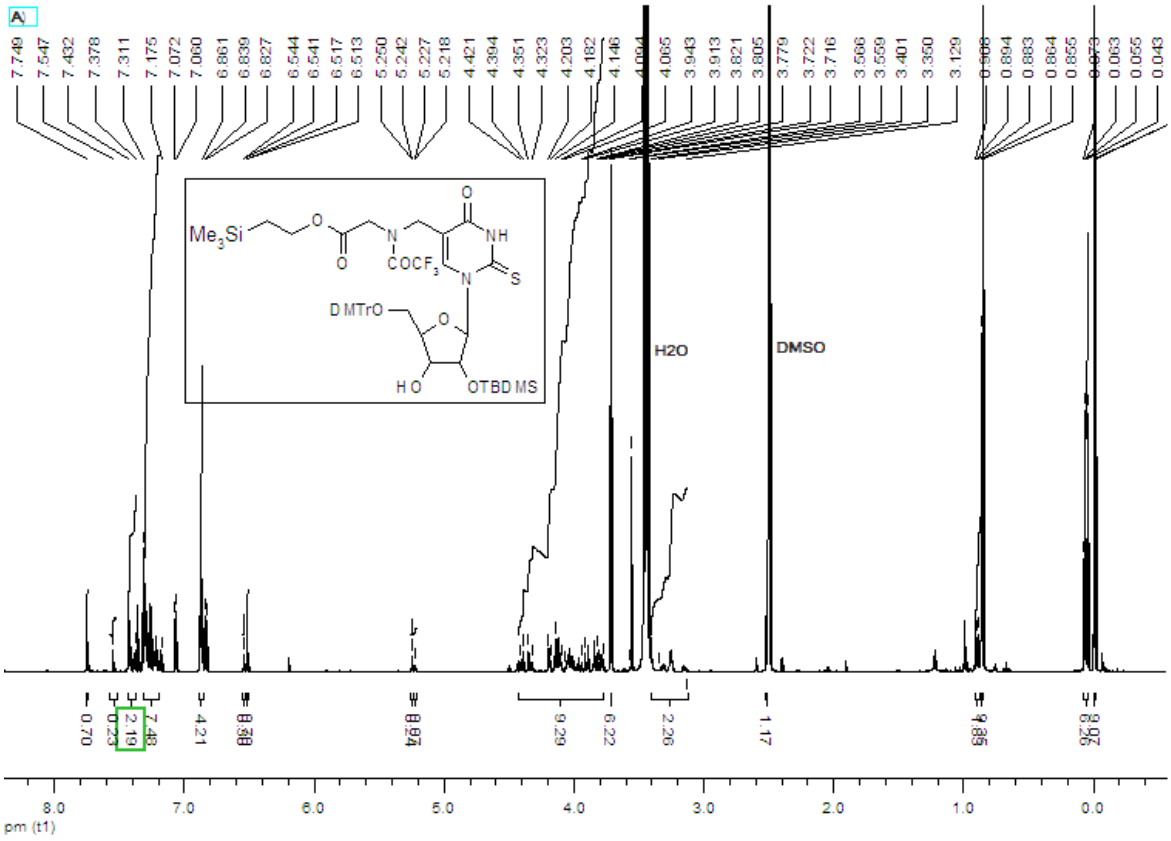
Supplementary Fig 3 A) ^{1}H NMR and COSY spectra of **12a** (700 MHz, $(CD_3)_2SO$); **B)** ^{13}C NMR spectrum of **12a** (176 MHz, C_6D_6).

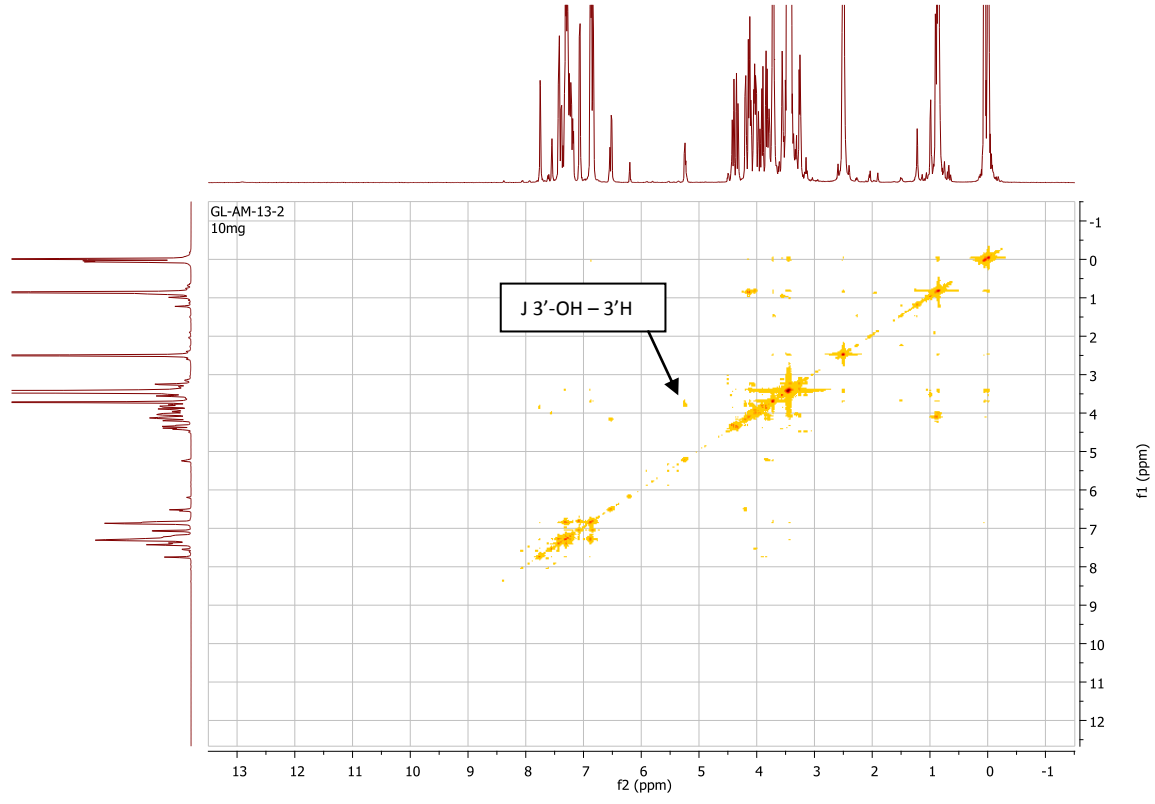


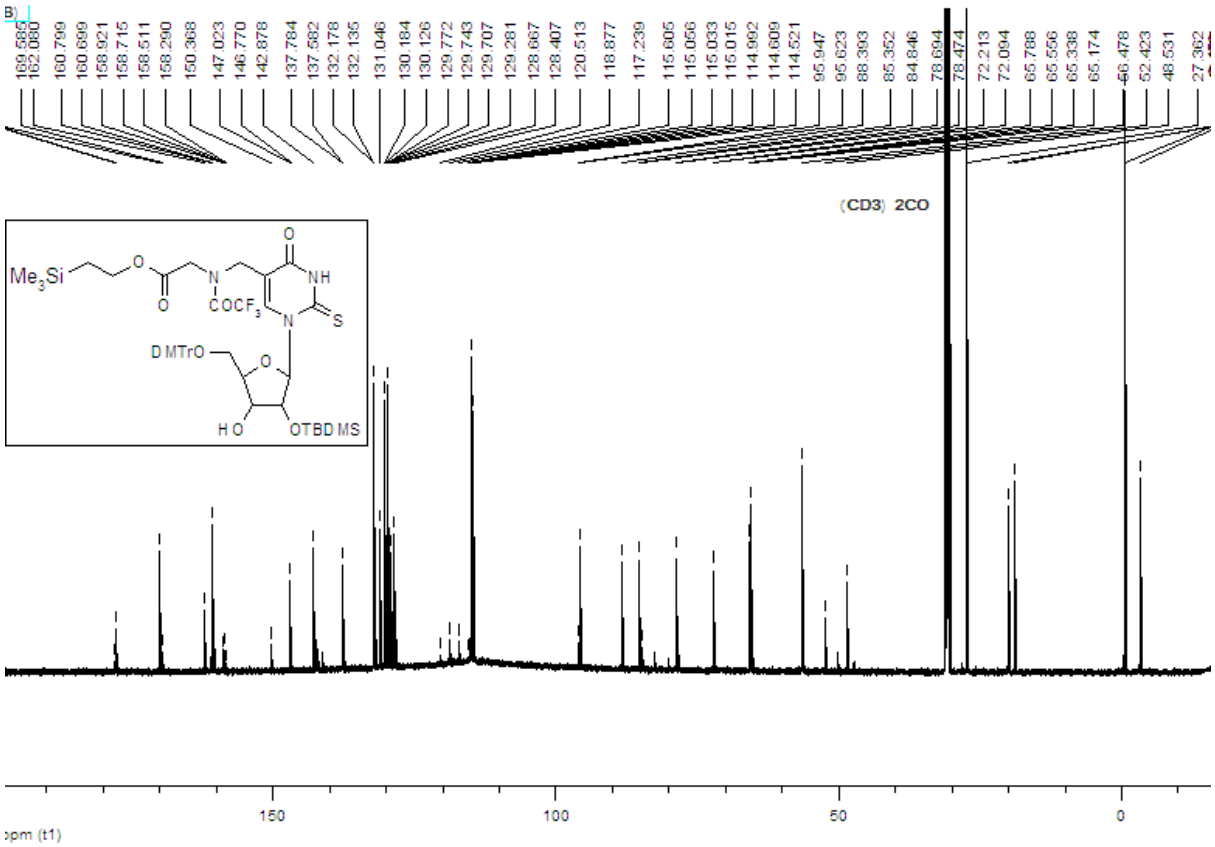

GL-AM-12c
15mg
J 3'-OH – 3'H
f2 (ppm)
f1 (ppm)


B)
C6D6
Me3Si
COCF3
DMTrO
HO
OTBDMS
ppm (t1)

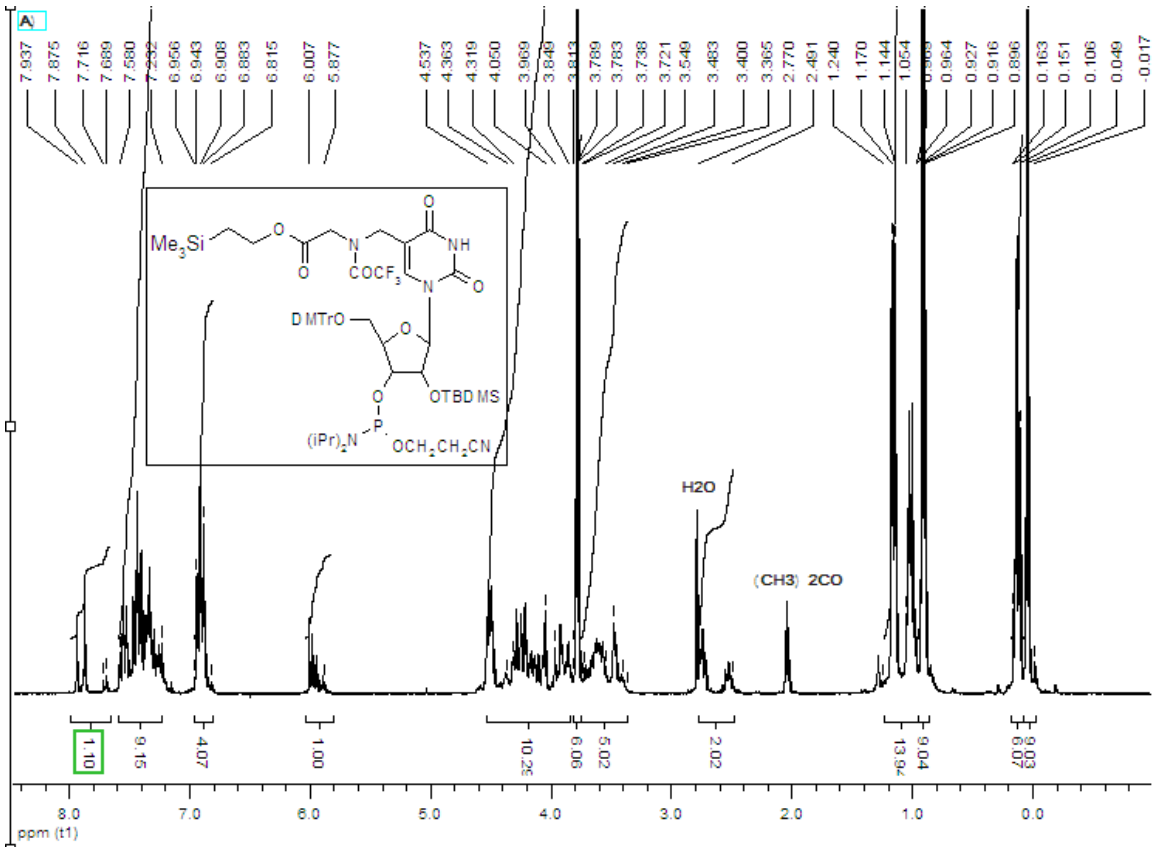
Supplementary Fig 4 A) ^{1}H NMR and COSY spectra of **12b** (700 MHz, $(CD_3)_2SO$); **B)** ^{13}C NMR spectrum of **12b** (176 MHz, $(CD_3)_2CO$).

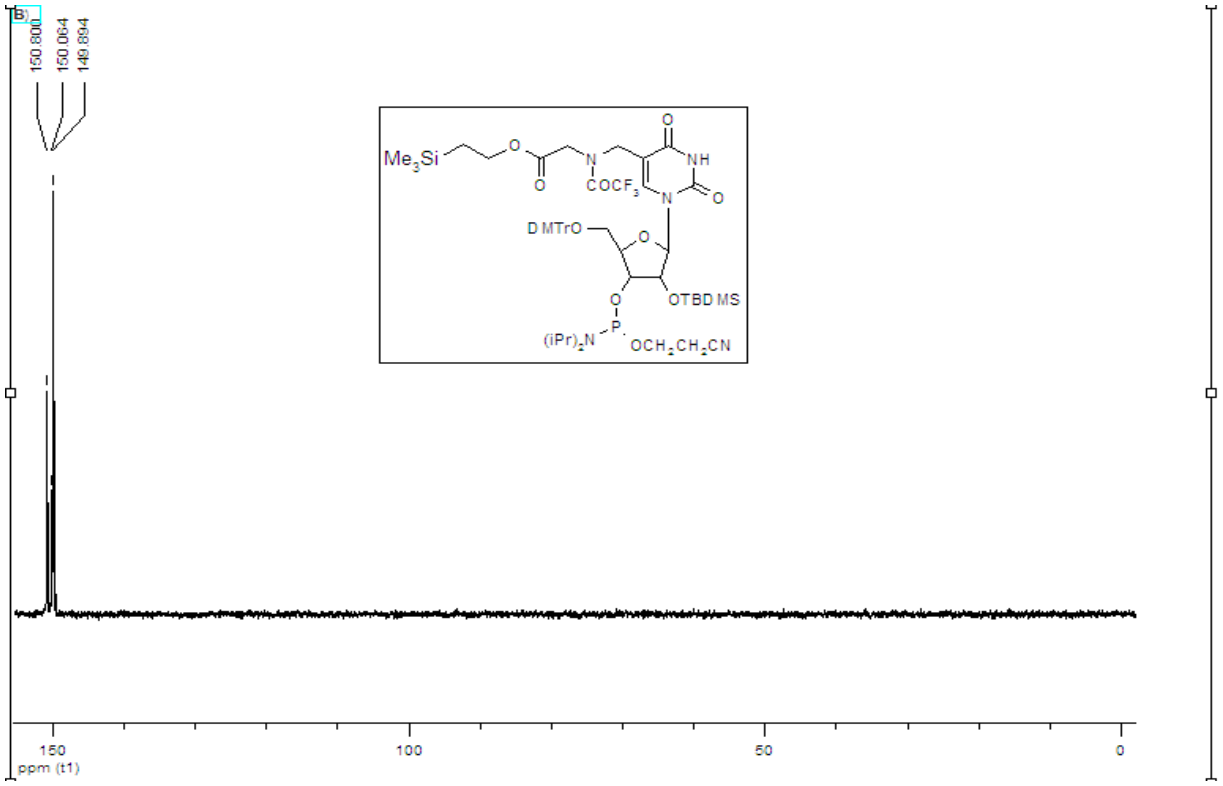




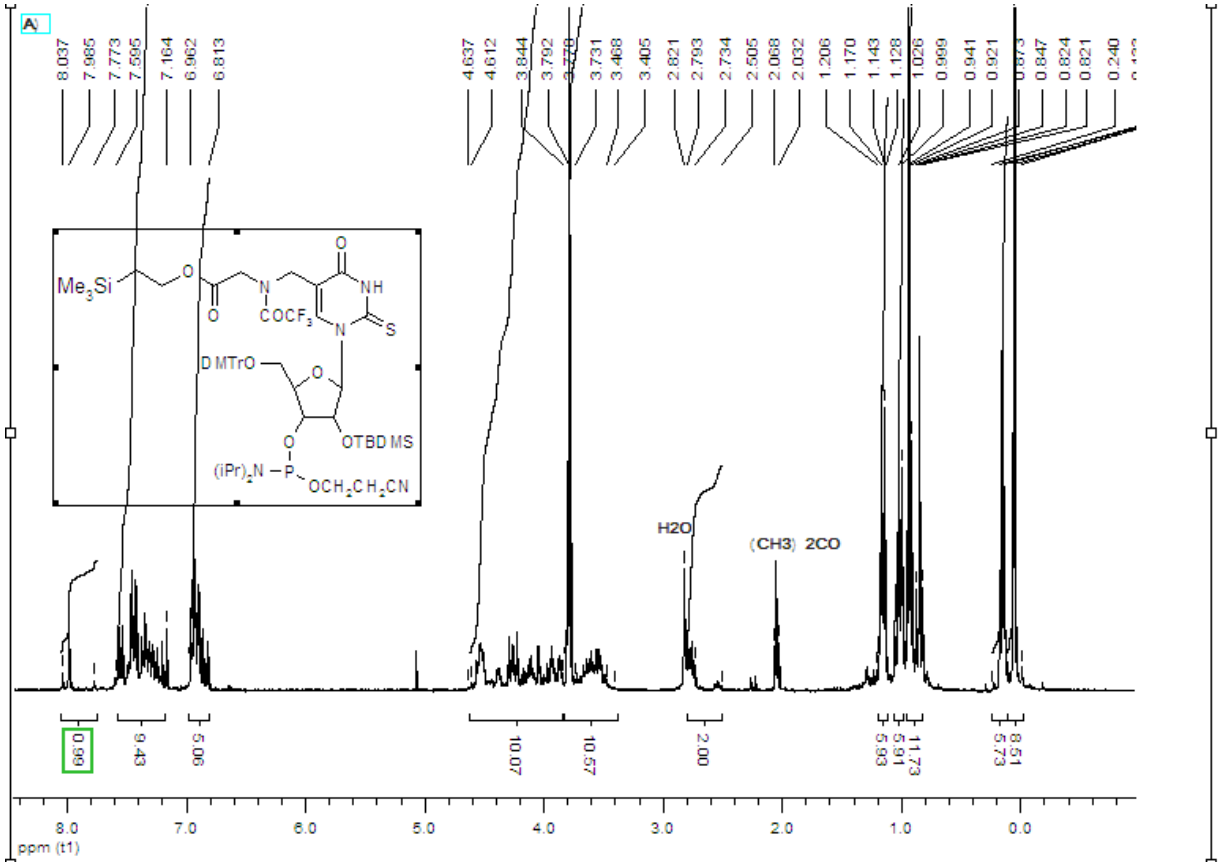


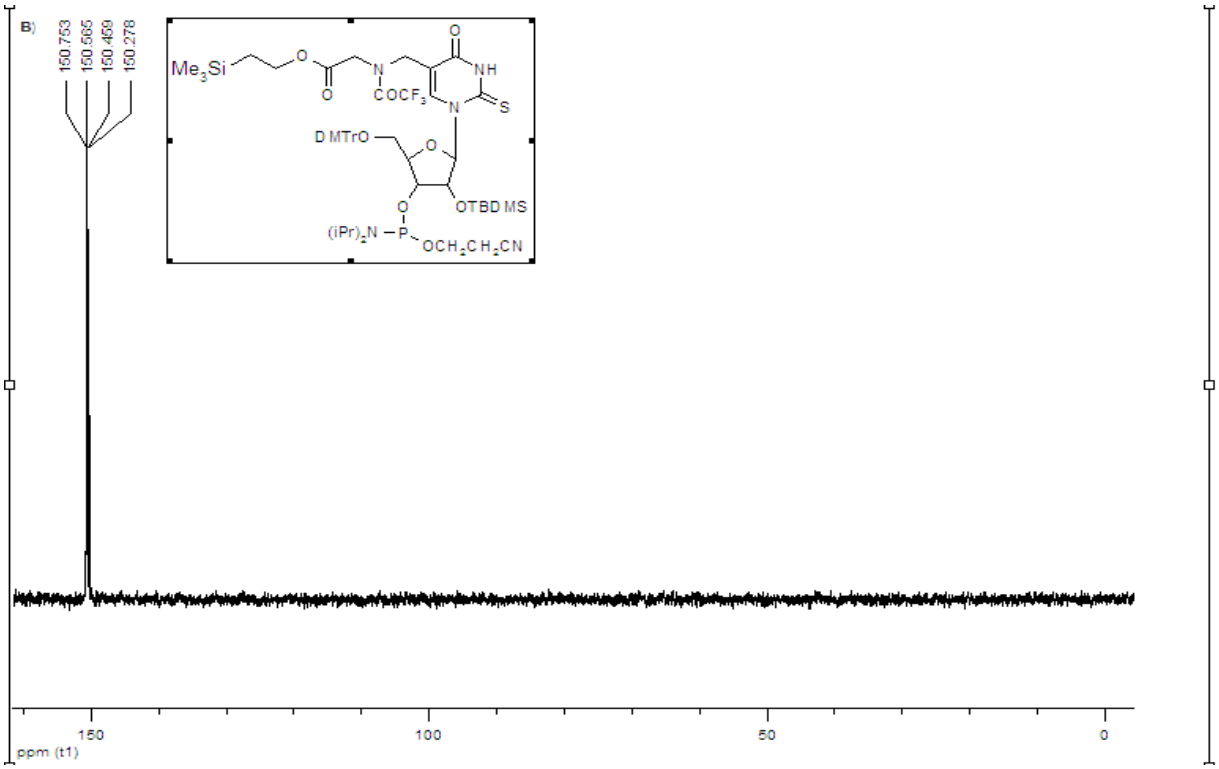
Supplementary Fig 5 A) ^{1}H NMR spectrum of **13a** (250 MHz, $(CD_3)_2CO$); **B)** ^{31}P NMR spectrum of **13a** (101 MHz, $(CD_3)_2CO$).





Supplementary Fig 6 A) ^{1}H NMR spectrum of **13b** (250 MHz, $(CD_3)_2CO$); **B)** ^{31}P NMR spectrum of **13b** (101 MHz, $(CD_3)_2CO$).





4. Oligoribonucleotide synthesis

Oligoribonucleotides were synthesized manually on a 5 μmole scale.The commercially available monomeric units A, C, U, and G were protected with DMT and TBS on the 5'- and 2'-hydroxy functions, respectively, and the exocyclic amine functions of A, C, and G were masked with 4-*tert*-butylphenoxyacetyl (tac) (Proligo®). Typical rC(tac)- or rU-succinyl-CPG (Proligo®) supports and 0.1 M acetonitrile solution of monomeric units were used. A, U, C, and G amidites were coupled in 8 molar excess for 8 min in the presence of Activator 42® (0.25 M solution of 5-(3,5-bis(trifluoromethyl)phenyl)-1*H*-tetrazole in ACN), while modified units were used in 12 molar excess and coupled twice, each time using 6 molar excess of amidite and 12 min coupling time. Capping was performed with tac anhydride (Fast protection Cap A : Cap B) for 2 min. 0.02 M I_2 oxidation solution in THF-H_2O-pyridine (8 eq) was used for 2 min.

Supplementary Table 1 Solid support synthesis protocol

Phosphoramidite condensation method		
Detritylation	3 % TCA/DCM	until the disappearance of the orange color of DMT^+ ion
Condensation	0.1 M phosphoramidite in ACN Activator 42®	8 min (canonical monomeric units) 2 × 12 min (modified monomeric units)
Capping	Cap A/Cap B 1.1/1 v/v	120 s
Oxidation	0.02 M iodine in THF/H_2O/py (90.54:9.05:0.41 v/v/v)	120 s

5. RNA deprotection and purification (for TMSE blockage of cmnm^5(s^2)U)

The "trityl-off" CPG-bound RNA was transferred from the column to a screw cap glass vial and 6.5 mL of TEA/ACN (1:1 v/v) was added. The solution was stirred for 25 min, and then the solvent was removed. The support-bound RNA was washed with ACN, dried in vacuo for 30 min and treated with 8.5 mL of 8 M ethanolic ammonia at rt for 8 h. The supernatant was removed and the support material was washed with an additional 3 mL of anhydrous ethanol. The combined washings were evaporated on a Speed-Vac concentrator.

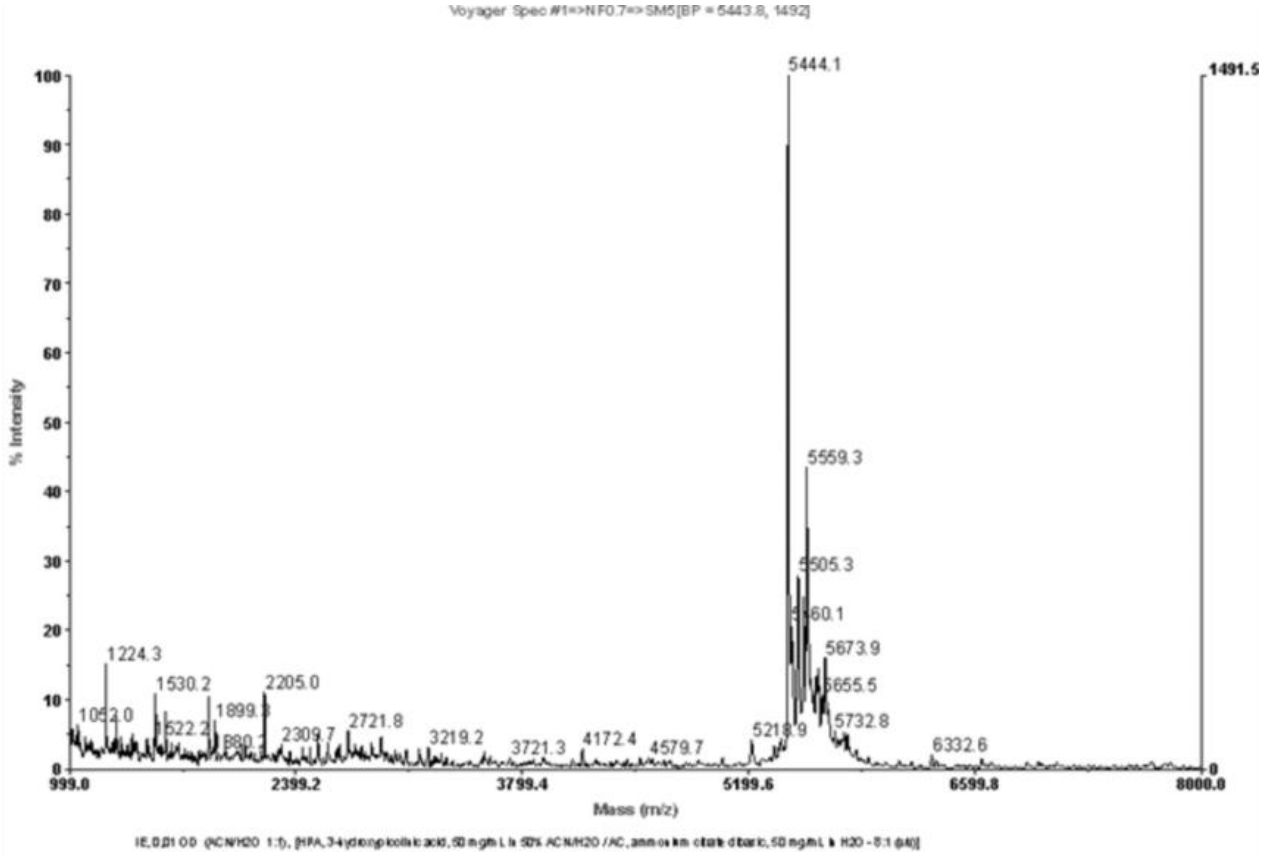
In the case of mt-ASL^{Leu}(cmnm^{5}U$_{34}$), the dried material was dissolved in 1.6 mL of 1 M tetrabutylammonium fluoride in *N*-methyl-2-pyrrolidinone (NMP), while mt-ASL^{Lys}(cmnm^{5}s^{2}U$_{34}$) was dissolved in 5 mL of 1 M tetraethylammonium fluoride in NMP and mixed. Both reactions were conducted at rt for 24 h, and then quenched by the addition of 0.05 M Na_2HPO_4-NaH_2PO_4 buffer solution (pH 7.6). Crude RNAs were desalted on a column packed with Sephadex® G-25 (elution with 20% aqueous ethanol), monitored by UV detection at 260 nm. The RNA-containing eluate was lyophilized and then purified by anion-exchange HPLC (Waters AP-2 column packed with TSK SuperQ®-5PW resin; elution with a linear gradient of NaBr (50–650 mM) in sterile 20 mM Na_2HPO_4-NaH_2PO_4 buffer solution (pH 7.5), containing EDTA (50 µM) and 10% ACN; flow 9 mL/min). Fractions containing the desired product were collected, concentrated, and desalted on a column packed with Sephadex® G-25. The desalted RNAs were lyophilized and analyzed.

6. RNA digestion and HPLC condition

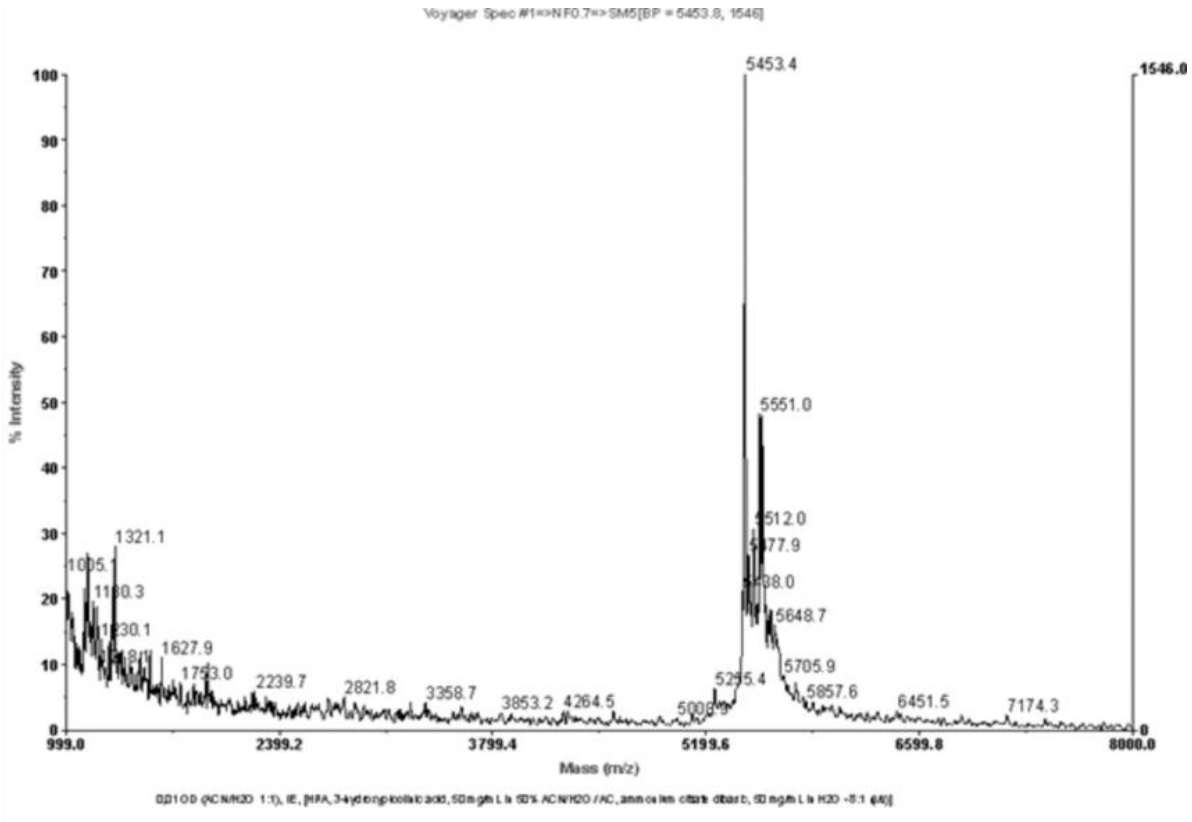
The 17-mer mt-ASL^{Leu}(cmnm^{5}U$_{34}$) or mt-ASL^{Lys}(cmnm^{5}s^{2}U$_{34}$) was digested into nucleosides with nuclease P_1 (Sigma, N8630) and alkaline phosphatase (Sigma, P4252) using standard protocol.[2] The resulting nucleoside mixtures were analyzed with a C18 separation column (ODS2, 4.6 mm × 250 mm) with a linear gradient of buffer A (10 mM KH_2PO_4; pH 5.3) and buffer B (20 % methanol in 10 mM KH_2PO_4; pH 5.1) with a flow 0.75 mL/min. The peaks were compared with reference samples of modified units in separate control experiments.

7. MALDI MS spectra of mt-ASL$^{Leu}_{S.cerevisiae}$ and mt-ASL$^{Lys}_{S.cerevisiae}$

Supplementary Fig 7 MALDI-TOF spectrum of mt-ASL$^{Leu}_{S.cerevisiae}$ modified with cmnm^{5}U. Experimental monoisotopic mass for $[M+H]^+$ is 5444.1, the calculated monoisotopic mass is 5444.3 Da.



Supplementary Fig 8 MALDI-TOF spectrum of mt-ASL$^{Lys}_{S.cerevisiae}$ modified with cmnm^{5}s^{2}U. Experimental monoisotopic mass for $[M+H]^+$ is 5453.4, the calculated monoisotopic mass is 5452.4 Da.



7. RNA deprotection (for NPE blockage of cmnm^5(s^2)U)

The "trityl-off" polymer-bound oligomers were treated with Et_3N/CH_3CN (1:1 v/v, 20 min, rt) to remove the β-cyanoethyl groups. The supernatant was removed and the CPG-oligomers were washed twice with acetonitrile. Next, 10% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in CH_3CN was added (40 minutes, 45 °C) for carboxyl deprotection. The supernatant was removed and CPG-oligomers were treated with 8 M NH_3/EtOH (8 h, rt) to remove –tac,-$COCF_3$, and to cleave the RNA from the CPG. The supernatant was evaporated and Et_3N•3HF (24 h, rt) was added for desilylation. The reaction was quenched by addition of ethoxytrimethylsilane and the crude RNA was precipitated. The RNA was collected by centrifugation, washed with *t*-butyl methyl ether and purified by IE-HPLC and desalted.

8. References

[1] (a) K. H. Scheit, *Chem. Ber.*, 1966, **99**, 3884; (b) K. Ikeda, S. Tanaka, Y. Mizuno, *Chem. Pharm. Bull.*, 1975, **23**, 2958.

[2] (a) C. W. Gehrke, K. C. Kuo, R. A. McCune, K. O. Gerhardt, P. F. Agris, *J. Chromatogr.*, 1982, **230**, 297; (b) C. W. Gehrke, K. C. Kuo, *J. Chromatogr.*, 1989, **471**, 3.