Supporting Information

# Base Pairing, Structural and Functional Insights into $N^4$-methylcytidine ($m^4C$) and $N^4,N^4$-dimethylcytidine ($m^4{}_2C$) Modified RNA

Song Mao,[1,2,#] Bartosz Sekula,[3,#] Milosz Ruszkowski,[4] Srivathsan V. Ranganathan,[2] Phensinee Haruehanroengra,[1,2] Ying Wu,[1,2] Fusheng Shen[1,2] and Jia Sheng[1,2,*]

[1] Department of Chemistry and [2] The RNA Institute, University at Albany, State University of New York, 1400 Washington Ave. Albany, NY, 12222, USA.
[3]Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Argonne, IL, USA
[4] Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
[#] Authors have equal contribution to this work
* Email: jsheng@albany.edu

**Table of Contents**

## Synthesis of m$^4$C phosphoramidite

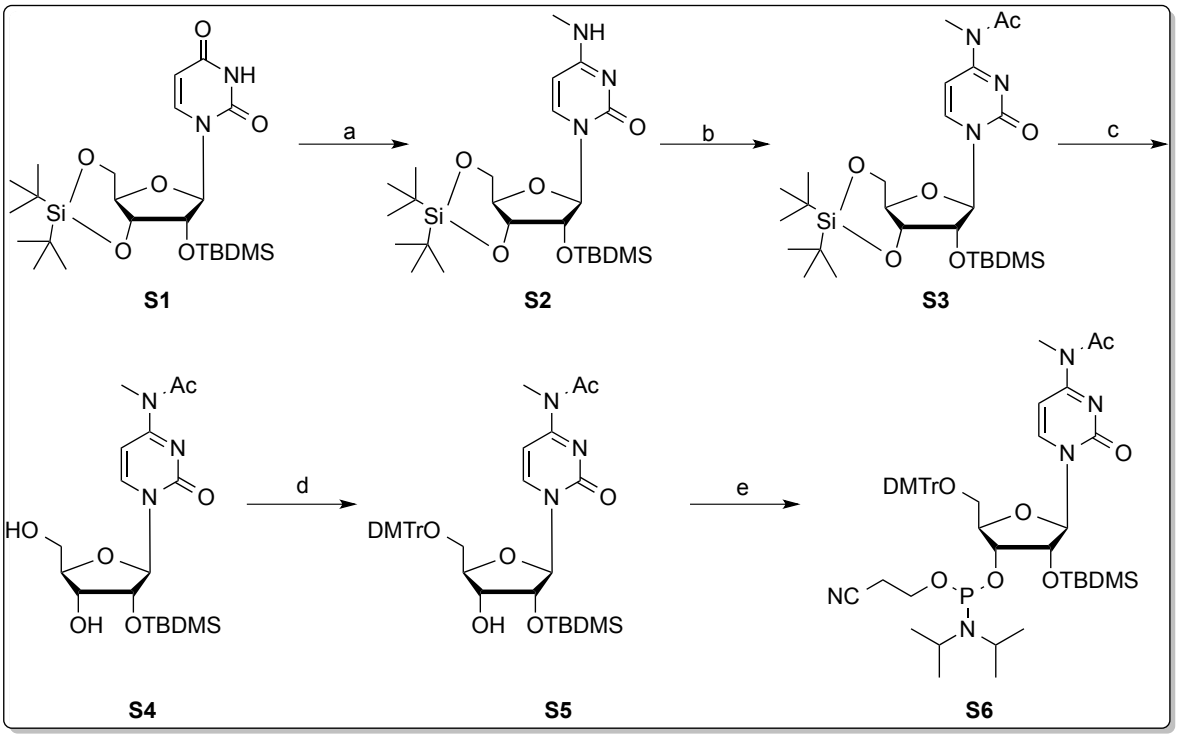


**Fig. S1**. Synthesis of m$^4$C phosphoramidite **S6**. Reagents and conditions: (a) 1) 2,4,6-triisopropylbenzenesulfonyl chloride, $Et_3N$, DMAP, $CH_2Cl_2$; 2) $MeNH_2$ (40% v/v aqueous solution), THF; (b) $Ac_2O$, $Et_3N$, DMAP, DCM (c) HF·Py, THF; (d) DMTrCl, Py; (e) (*i*-$Pr_2N)_2P(Cl)OCH_2CH_2CN$, (*i*-$Pr)_2NEt$, 1-methylimidazole, DCM.

1-(2′-*O*-*tert*-butyldimethylsilyl-3′,5′-*O*-di-*tert*-butylsilylene-beta-D-ribofuranosyl)-*N*$^4$-methylcytidine **S2**.

To a solution of compound **S1** (996 mg, 2 mmol), $Et_3N$ (0.66 mL, 4 mmol) and DMAP (24 mg, 0.1 mmol) in DCM (20 mL) was added 2,4,6-triisopropylbenzenesulfonyl chloride (TPSCl, 788 mg, 2.6 mmol) at 0°C under Ar. The resulting solution was stirred at room temperature for 3 h. TLC showed that the starting material was consumed up. The mixture was diluted with DCM (100 mL) and washed with sat. $NaHCO_3$ and brine. The organic layer was dried by $Na_2SO_4$ and concentrated under vacuum to obtained yellow solid. Then the yellow solid was dissolved in THF (20 mL), $MeNH_2$ (10 mL, 40% v/v aqueous solution) was added. The mixture was stirred at r.t. for 15 h. The solvent was removed and the residue was purified by silica gel chromatography to give compound **S2** (640 mg, 1.2 mmol, 60% yield) as a white solid. TLC $R_f$ = 0.4 (DCM:MeOH = 20:1). $^1$H NMR (500 MHz, $CDCl_3$) δ 7.28 (d, *J* = 7.5 Hz, 1H), 5.66 (s, 1H), 5.56 (s, *J* = 7.0 Hz, 1H), 4.53-4.49 (m, 1H), 4.37-4.35 (m, 1H), 4.25-4.19 (m, 1H), 4.00-3.95 (t, *J* = 10.0 Hz, 1H), 3.87-3.83 (dd, *J* = 4.5, 9.5 Hz, 1H), 3.01 (d, *J* = 4.5 Hz, 3H), 1.03 (s, 9H), 1.02 (s, 9H), 0.94 (s, 9H), 0.25 (s, 3H), 0.16 (s, 3H).

$N^4$-Acetyl-1-(2′-*O*-*tert*-butyldimethylsilyl-3′,5′-*O*-di-*tert*-butylsilylene-beta-D-ribofuranosyl)-$N^4$-methylcytidine **S3**.

To a solution of compound **S2** (640 mg, 1.2 mmol), DIPEA (0.8 mL, 4.8 mmol) and DMAP (13 mg, 0.12 mmol) in DCM (15 mL) was added $Ac_2O$ (0.3 mL, 2.4 mmol) under Ar. The resulting mixture was stirred at room temperature for 20 h. The solution was diluted with DCM (100 mL), washed with saturated $NaHCO_3$ and brine. The organic layer was dried by $Na_2SO_4$, filtered and evaporated under reduced pressure. The residue was purified by silica gel chromatography to give compound **S3** (600 mg, 1.0 mmol, 90% yield) as a white solid. TLC $R_f$ = 0.7 (DCM:MeOH = 20:1). $^1$H NMR (500 MHz, $CDCl_3$) δ 7.63 (d, *J* = 8.0 Hz, 1H), 7.20 (d, *J* = 8.0 Hz, 1H), 5.71 (s, 1H), 4.55 (dd, *J* = 5.0, 9.0 Hz, 1H), 4.36 (d, *J* = 4.0 Hz, 1H), 4.32-4.26 (m, 1H), 4.00 (dd, *J* = 9.5, 10.5 Hz, 1H), 3.81 (dd, *J* = 4.5, 9.5 Hz), 3.48 (s, 3H), 2.41 (s, 3H), 1.03 (s, 9H), 10.2 (s, 9H), 0.96 (s, 9H), 0.27 (s, 3H), 0.18 (s, 3H).

$N^4$-Acetyl-1-(2′-*O*-*tert*-butyldimethylsilyl-beta-D-ribofuranosyl)-$N^4$-methylcytidine **S4**.

To a solution of compound **S3** (280 mg, 0.5 mmol) in THF (5 mL) at 0 °C was added a solution of hydrogen fluoride-pyridine complex (hydrogen fluoride ~70%, pyridine ~30%; 0.1 mL) in pyridine (0.6 mL). After 1 h at 0 °C the reaction was complete and pyridine (2 mL) was added. The reaction mixture was washed with sat. $NaHCO_3$, dried over $Na_2SO_4$ and evaporated. The residue was purified by silica gel chromatography to give compound **S4** (150 mg, 0.36 mmol, 73% yield) as a white solid. TLC $R_f$ = 0.3 (DCM:MeOH = 20:1). $^1$H NMR (400 MHz, $CDCl_3$) δ 7.99 (d, *J* = 7.6 Hz, 1H), 7.19 (d, *J* = 7.2 Hz, 1H), 5.53 (s, 1H), 4.22-4.14 (m, 2H), 4.02-3.98 (m, 1H), 3.90 (m, 1H), 3.83-3.79 (m, 1H), 3.45 (s, 3H), 2.40 (s, 3H), 0.88 (s, 9H), 0.11-0.09 (d, 6H).

$N^4$-Acetyl-1-(2′-*O*-*tert*-butyldimethylsilyl-5′-*O*-4,4′-dimethoxytrityl-5′-beta-D-ribofuranosyl)-$N^4$-methylcytidine **S5**.

To a solution of compound **S4** (150 mg**,** 0.36 mmol) in dry pyridine (5 mL) was added 4,4′-dimethoxytrityl chloride (244 mg, 0.72 mmol) under Ar. The resulting solution was stirred at room temperature overnight. The reaction was quenched with methanol (1 mL) and stirred for another 5 min. The reaction mixture was then concentrated to dryness under vacuum. The residue was purified by silica gel chromatography to give compound **S5** (150 mg, 0.21 mmol, 58% yield) as a white solid. TLC $R_f$ = 0.4 (DCM:EA = 1:1). $^1$H NMR (400 MHz, $CDCl_3$) δ 8.41 (d, *J* = 7.6 Hz, 1H), 7.43-7.40 (m, 2H), 7.33-7.30 (m, 6H), 7.28-7.23 (m, 1H), 6.86 (d, *J* = 8.8 Hz, 1H), 6.65 (d, *J* = 7.6 Hz, 1H), 5.87 (s, 1H), 4.41-4.34 (m, 1H), 4.30-4.28 (m, 1H), 4.15-4.07 (m, 1H), 3.81 (s, 6H), 3.61-3.58 (m, 1H), 3.55-3.51 (m, 1H), 3.42 (s, 3H), 2.39 (s, 3H), 0.94 (s, 9H), 0.35 (s, 3H), 0.21 (s, 3H).

$N^4$-Acetyl-1-(2′-*O*-*tert*-butyldimethylsilyl-3′-*O*-(2-cyanoethyl-*N,N*-diisopropylamino) phosphoramidite-5′-*O*-4,4′-dimethoxytrityl-5′-beta-D-ribofuranosyl)-$N^4$-methylcytidine **S6**.

To a solution of compound **S5** (150 mg, 0.21 mmol) in DCM (3 mL) was added DIPEA (0.14 mL, 0.8 mmol), 1-methyl-1*H*-imidazole (17 μL, 0.21 mmol) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (0.1 mL, 0.4 mmol). The resulting solution was stirred at

room temperature overnight under Ar. The mixture was diluted with DCM (50 mL), washed with brine. The organic layer was dried by $Na_2SO_4$, filtered and evaporated under reduced pressure. The residue was purified by silica gel chromatography to give compound **S6** (140 mg, 0.15 mmol, 73% yield) as a white solid. TLC $R_f$ = 0.4 (DCM:EA = 1:1). $^1$H NMR (400 MHz, $CDCl_3$) δ 8.53-8.41 (m, 1H), 7.47-4.26 (m, 9H), 6.87-6.83 (m, 4H), 6.54-6.34 (m, 1H), 5.88-5.79 (m, 1H), 4.36-4.30 (m, 2H), 3.81-3.80 (d, 6H), 3.75-3.64 (m, 2H), 3.56-3.46 (m, 3H), 3.41-3.37 (d, 3H), 2.40-2.37 (d, 3H), 1.30-1.08 (m, 12H), 0.93-0.91 (d, 9H), 0.29 (s, 3H), 0.16 (s, 3H). $^{31}$P NMR (162 MHz, $CDCl_3$) δ 150.60, 148.76. HRMS (ESI-TOF) $[M+H]^+$ = 916.4374 (calc. 916.4446). Chemical formula: $C_{48}H_{66}N_5O_9PSi$.

## $^{1}$H and $^{13}$C NMR and HRMS spectra of synthesized compounds

## $^{1}$HNMR, $^{13}$CNMR and Mass Spectrum

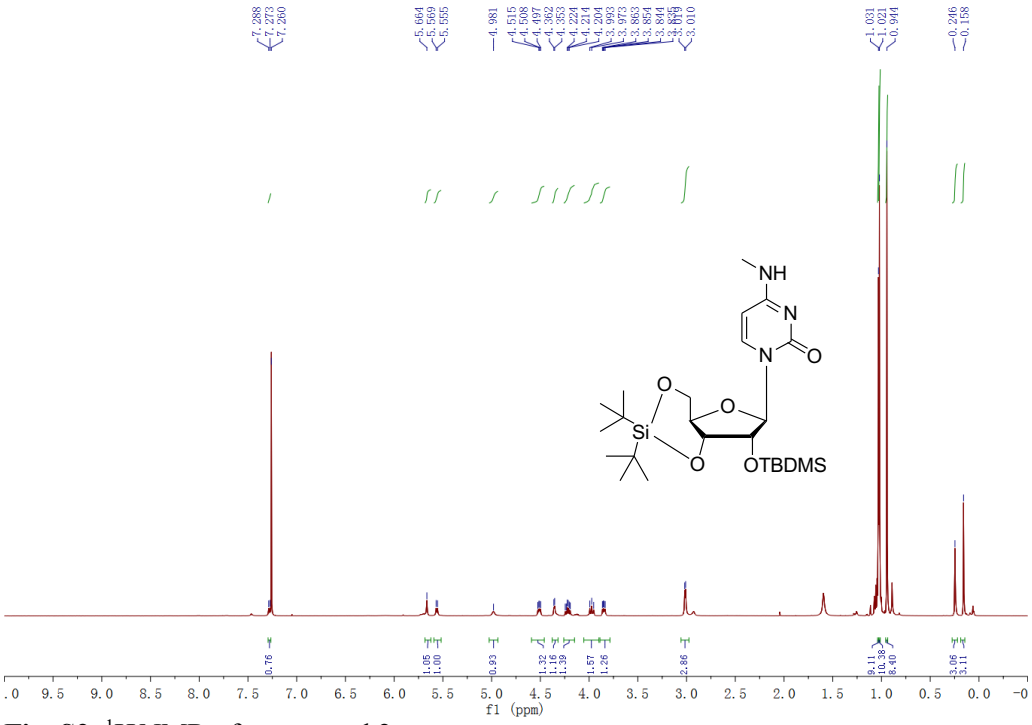


**Fig. S2.** $^{1}$H NMR of compound **2**.

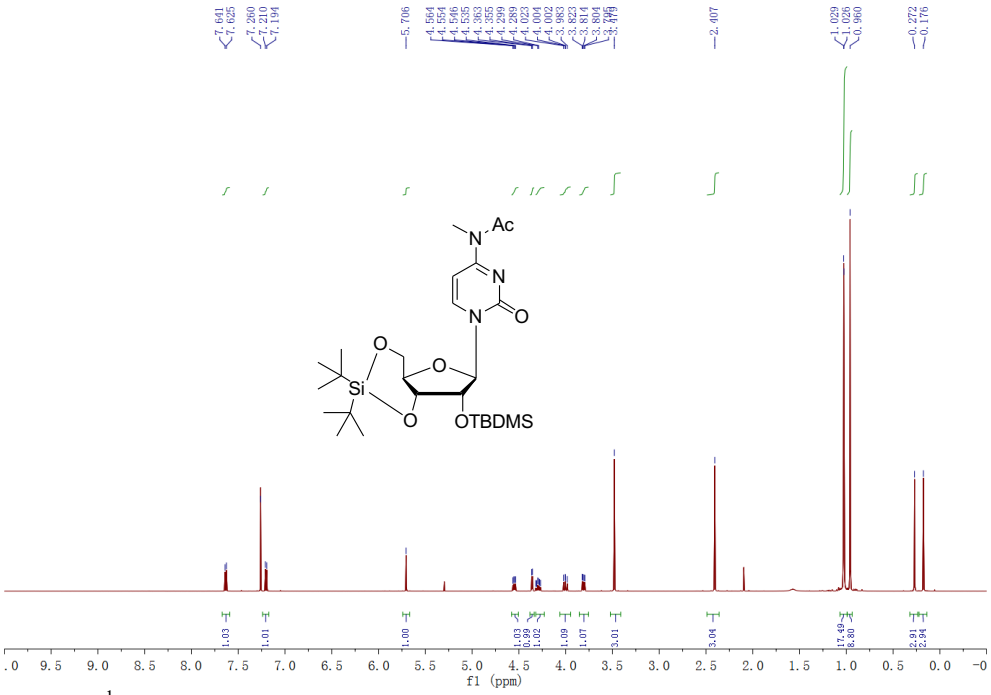


**Fig. S3.** $^{1}$H NMR of compound **3**.

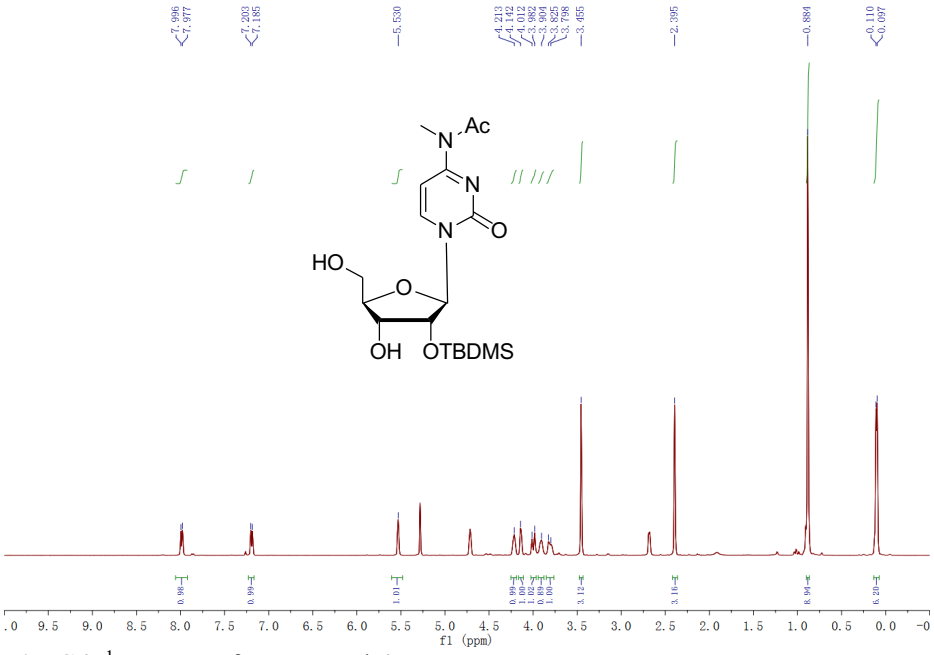


**Fig. S4.** $^{1}$H NMR of compound **4**.

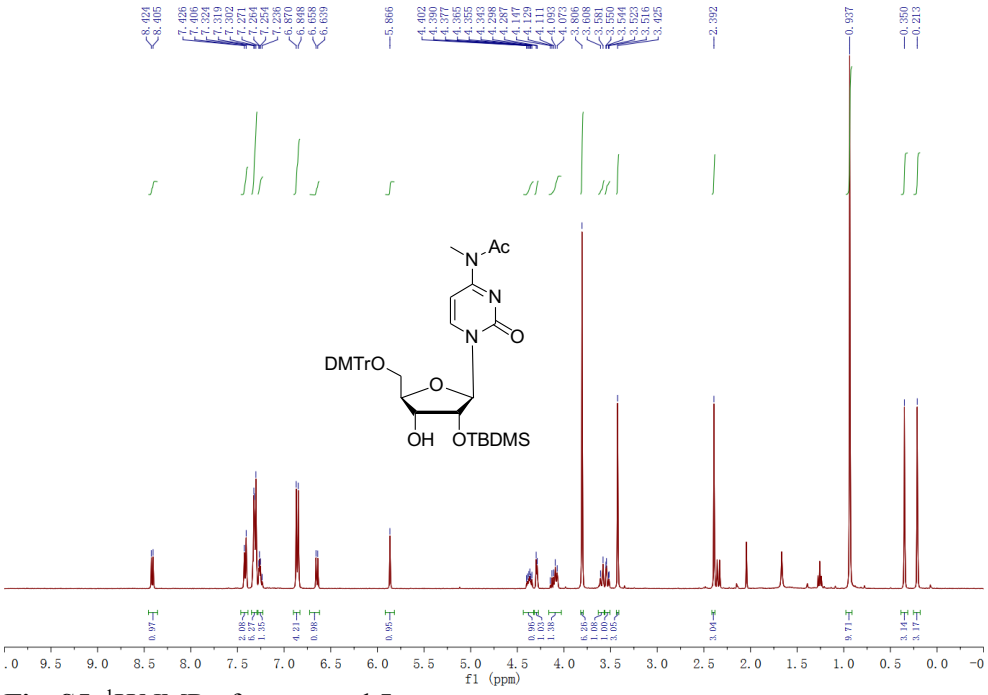


**Fig. S5.** $^{1}$H NMR of compound **5**.

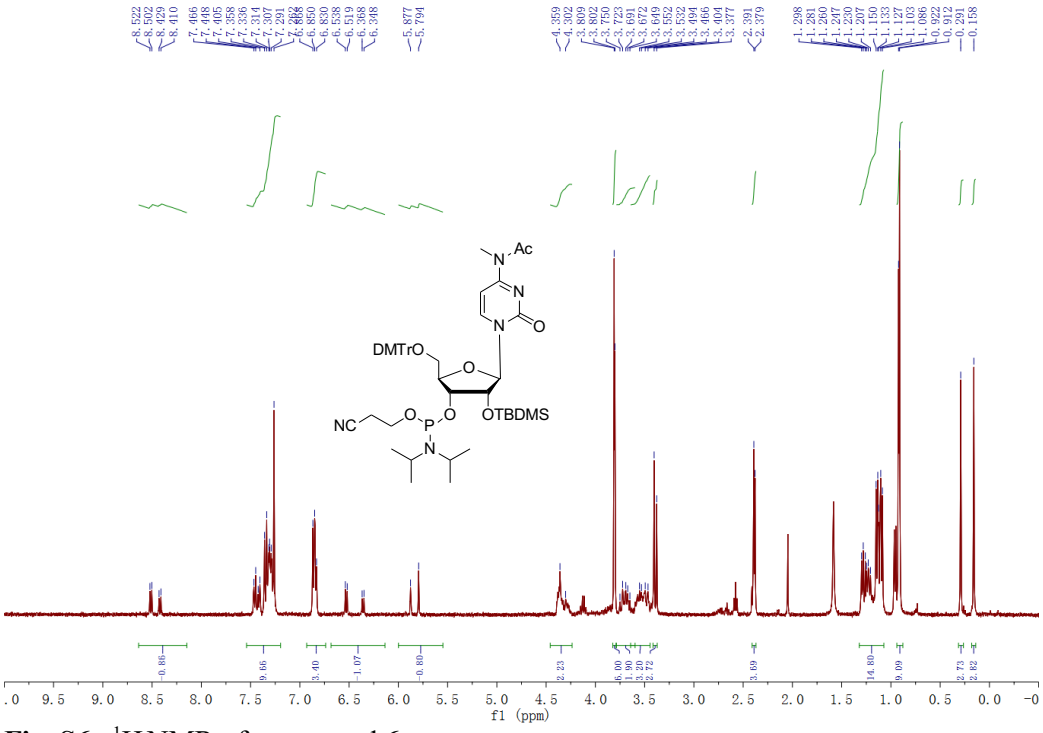


**Fig. S6 .** $^{1}$H NMR of compound **6**.

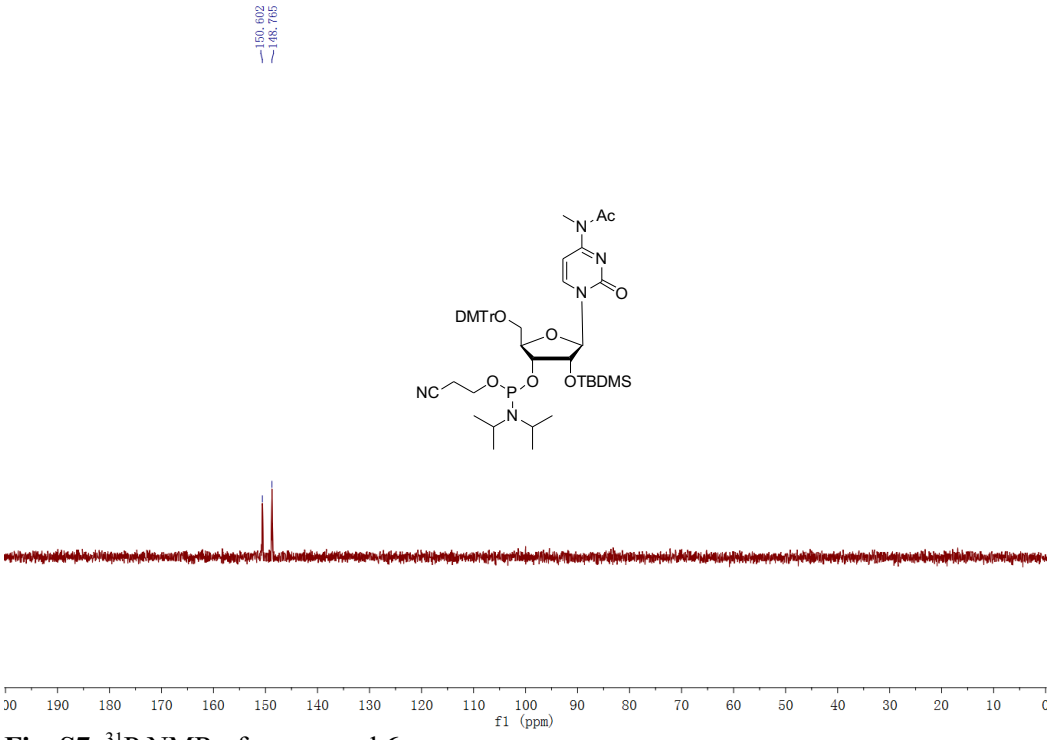


**Fig. S7.** $^{31}$P NMR of compound **6**.

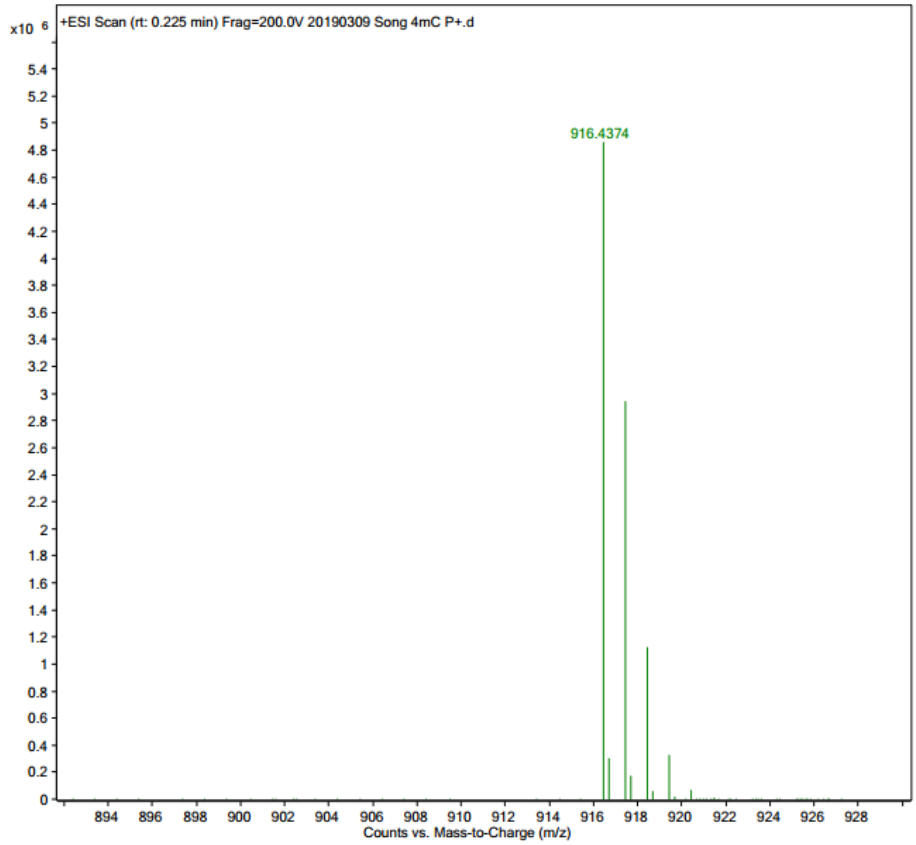


**Fig. S8.** Mass of compound **6**.

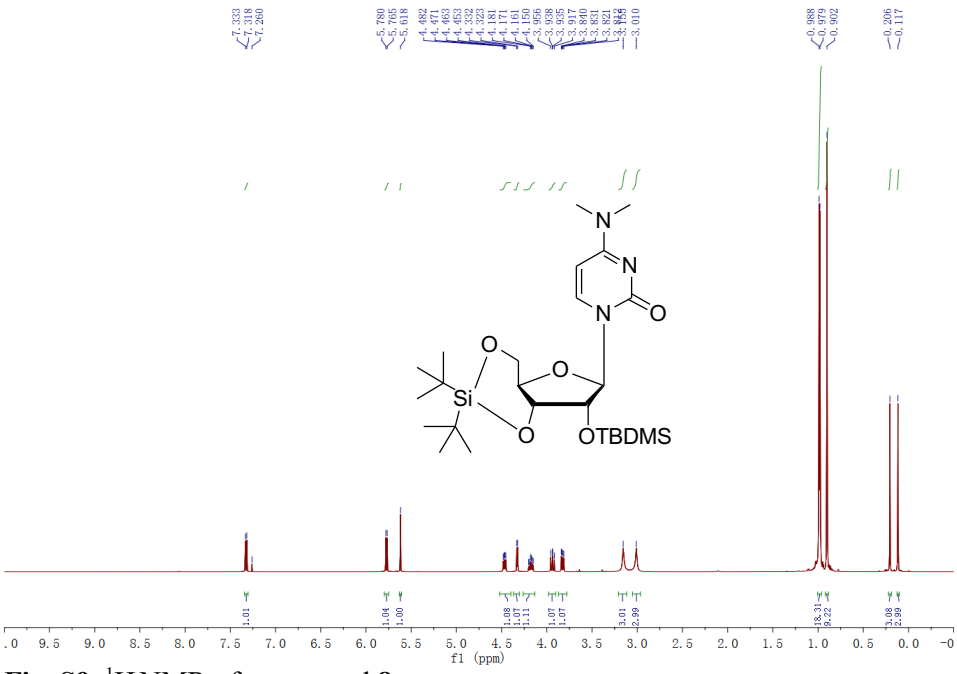


**Fig. S9.** $^{1}$H NMR of compound **8**.

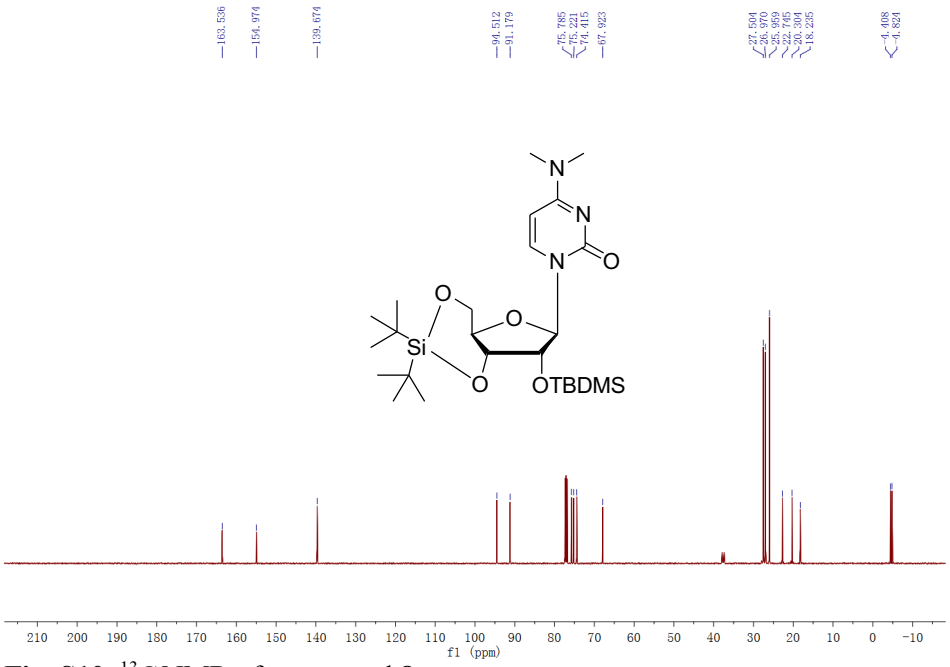


**Fig. S10.** $^{13}$C NMR of compound **8**.

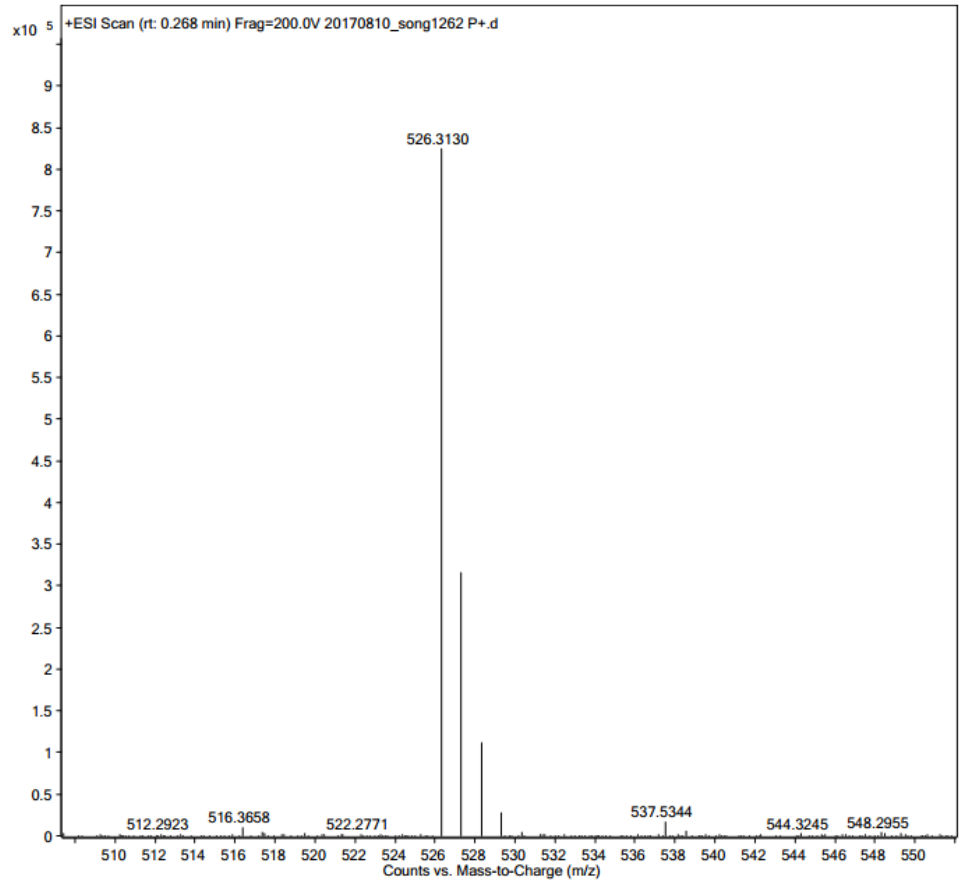


**Fig. S11.** Mass of compound **8**.

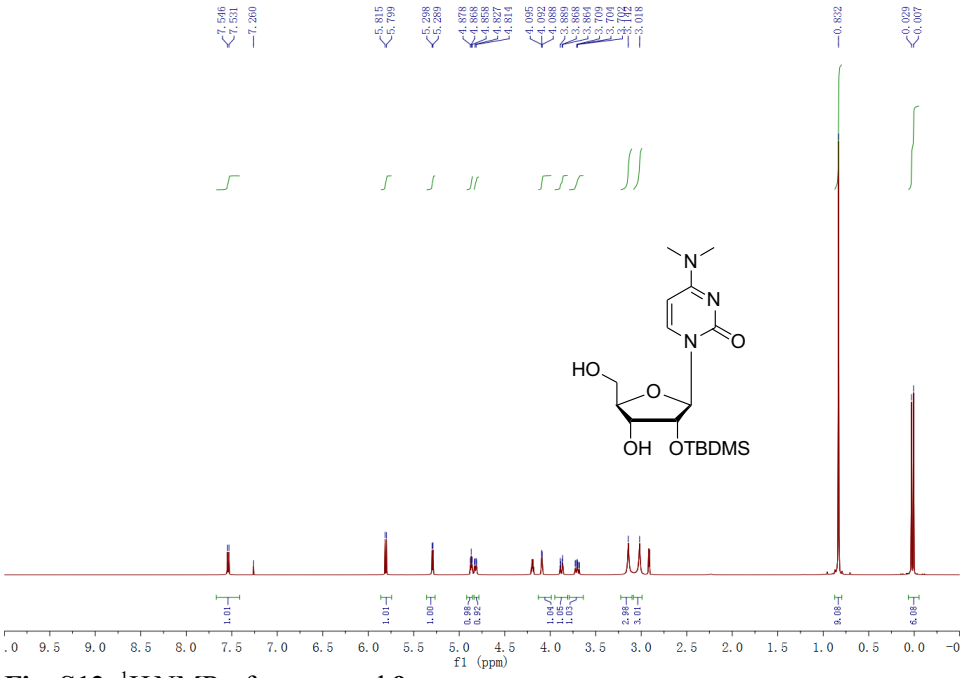


**Fig. S12.** $^{1}$H NMR of compound **9**.

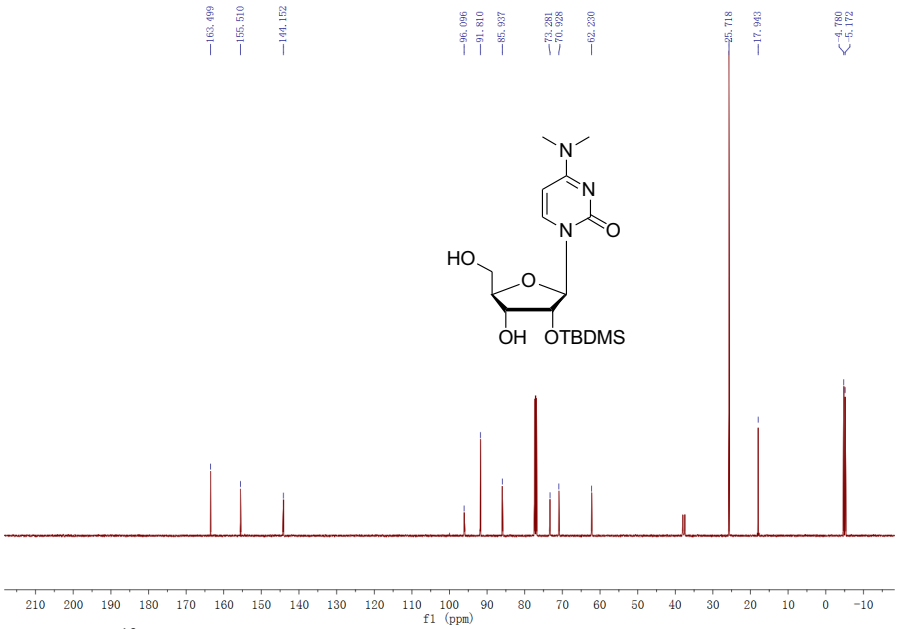


**Fig. S13.** $^{13}$C NMR of compound **9**.

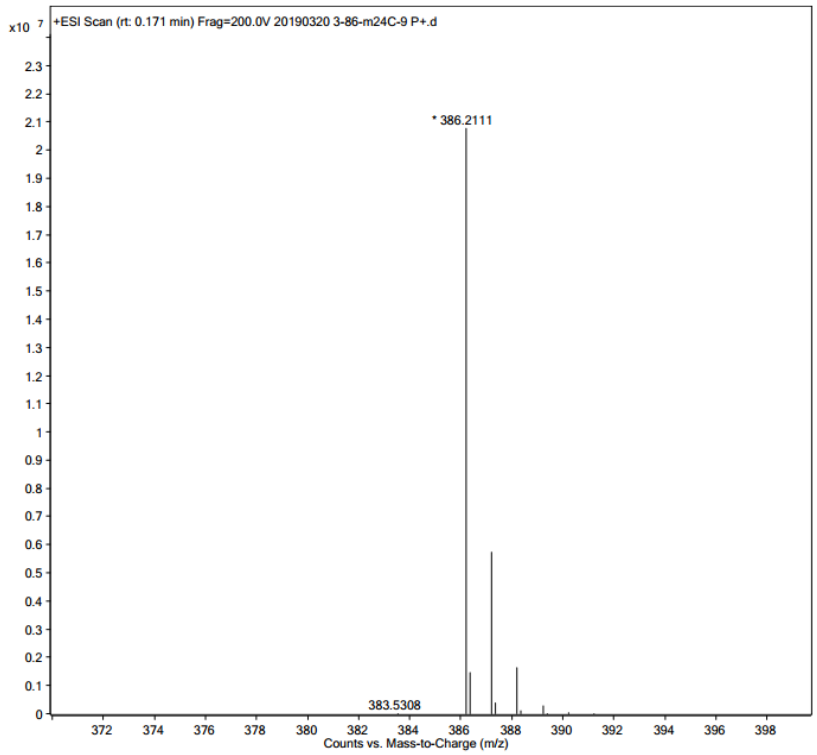


**Fig. S14.** Mass of compound **9**.

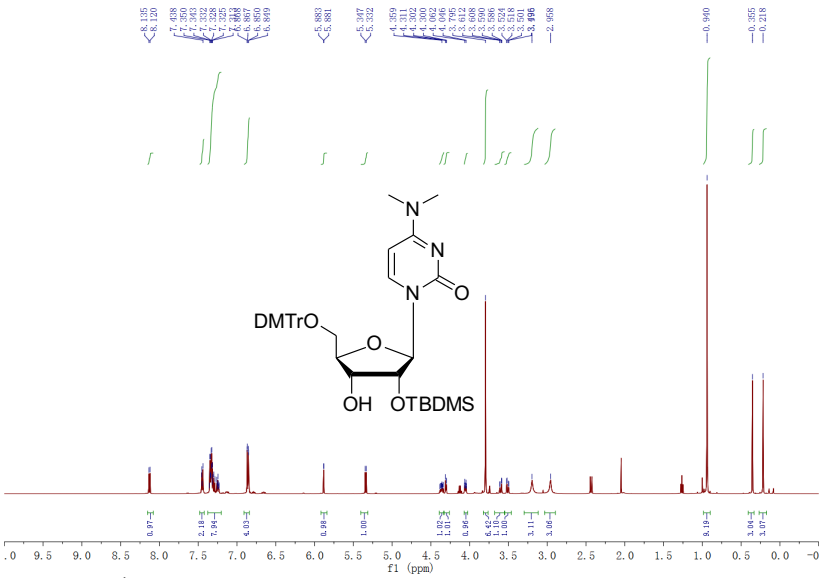


**Fig. S15.** $^{1}$H NMR of compound **10**.

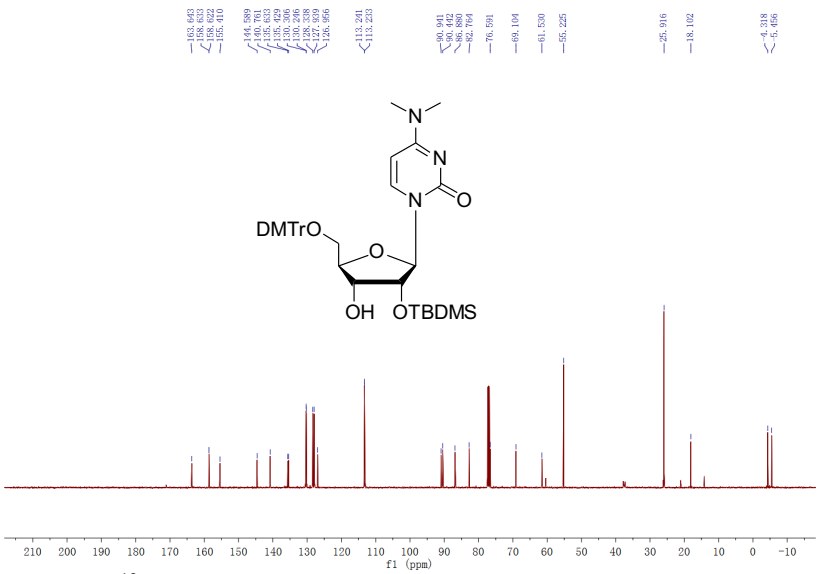


**Fig. S16.** $^{13}$C NMR of compound **10**.

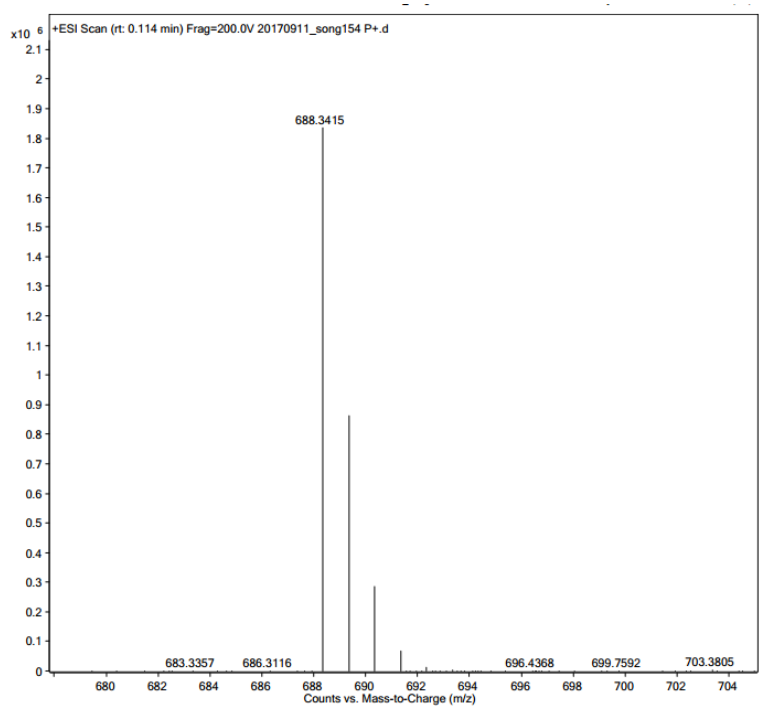


**Fig. S17.** Mass of compound **10**.

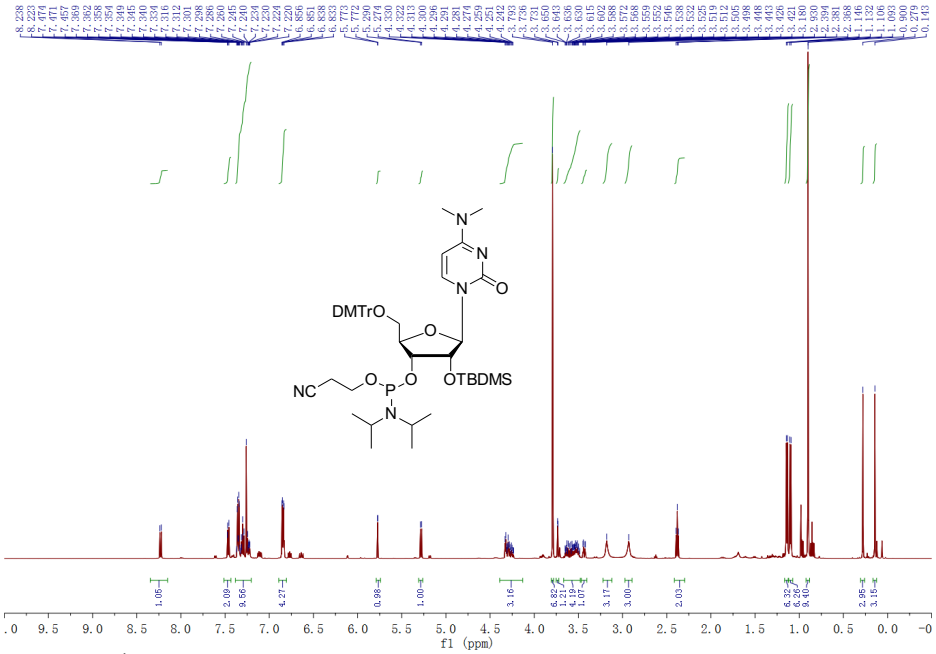


**Fig. S18.** $^{1}$H NMR of compound **11**.

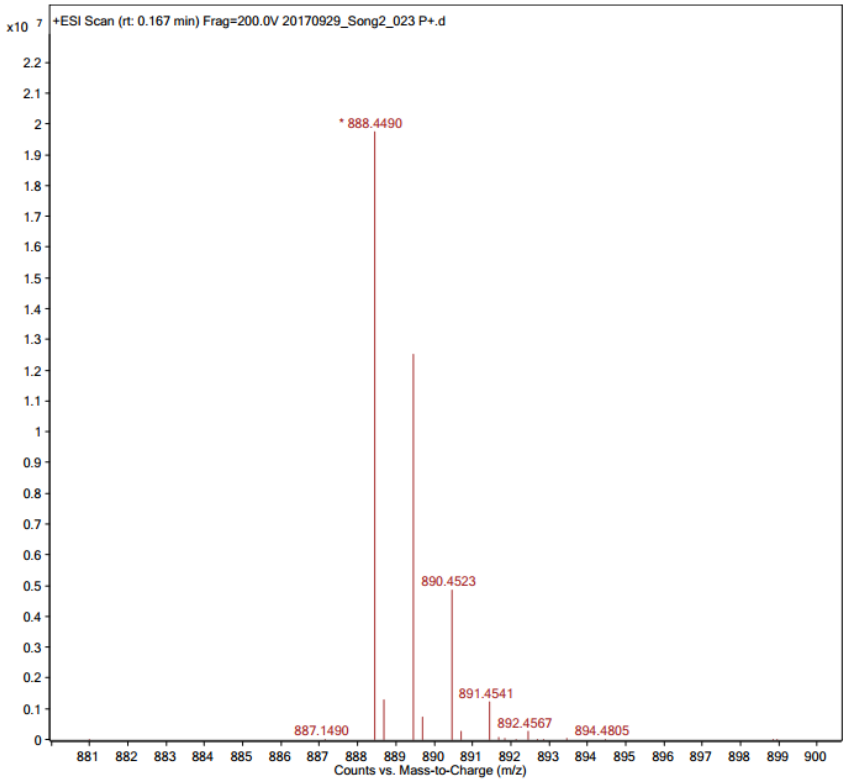


**Fig. S19.** Mass of compound **11**.

**Synthesis, HPLC and Characterization of RNA oligonucleotides**

**HPLC purification and analysis.**

The oligonucleotides were purified by IE-HPLC at a flow rate of 1 mL/min. Buffer A was 20 mM Tris-HCl, pH 8.0; buffer B 1.25M NaCl in 20 mM Tris-HCl, pH 8.0. A linear gradient from 100% buffer A to 70% buffer B in 20 min was used to elute the oligonucleotides. The analysis was carried out by using the same type of analytical column with the same eluent gradient. The 31-mer RNA templates were purified on a preparative 20% denaturing polyacrylamide gel (PAGE). All the modified-oligonucleotides were checked by MALDI-TOF, as summarized in Table S1 and Fig. S19-33.

**Table S1.** RNA sequences containing $m^4C$ and $m^4{}_2C$.

| Entry | RNA Sequences | Measured (calc.) m/z |
|---|---|---|
| **ON1** | 5′-AAUGC$m^4$CGCACUG-3′ | $[M+H]^+$ = 3807.3 (3807.6) |
| **ON2** | 5′-GGACU$m^4$CCUGCAG-3′ | $[M+H]^+$ = 3823.3 (3823.6) |
| **ON3** | 5′-CCGG$m^4$CGCCGG-3′ | $[M+H]^+$ = 3203.7 (3203.5) |
| **ON4** | 5′-AAUGC$m^4{}_2$CGCACUG-3′ | $[M+H]^+$ = 3821.4 (3821.6) |
| **ON5** | 5′-GGACU$m^4{}_2$CCUGCAG-3′ | $[M+H]^+$ = 3837.6 (3837.6) |
| **ON6** | 5′-CCGG$m^4{}_2$CGCCGG-3′ | $[M+H]^+$ = 3217.7 (3217.5) |
| **ON7** | 5′-U$m^4{}_2$CGUACGA-3′ | $[M+H]^+$ = 2537.3 (2537.4) |
| **ON8** | 5′-GUA$m^4{}_2$CGUAC-3′ | $[M+H]^+$ = 2537.9 (2537.4) |

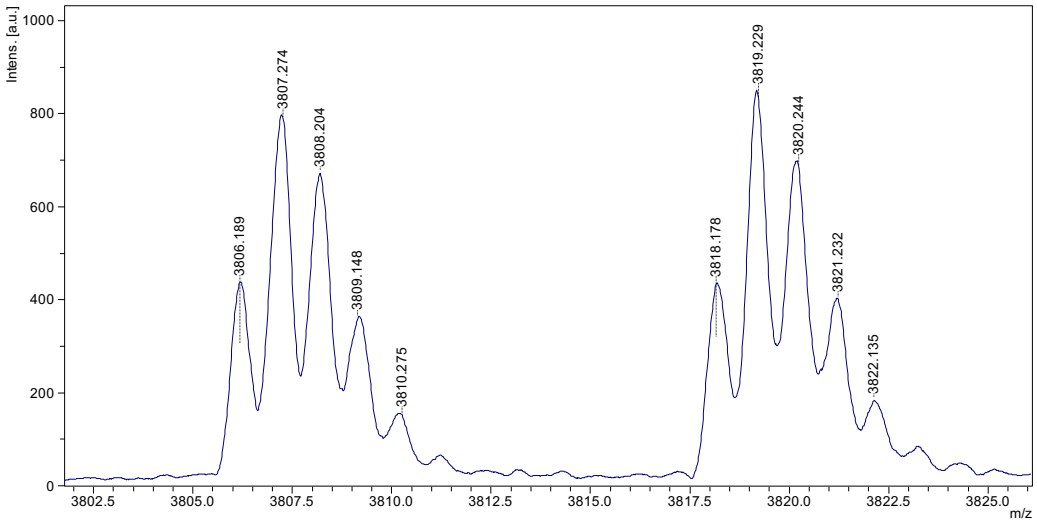


**Figure S20.** MALDI-TOF MS of **ON1** $m^4$C-12 mer (5′-AAUGC$m^4$CGCACUG-3′) $[M+H]^+$ = 3807.3 (calc. 3807.6).

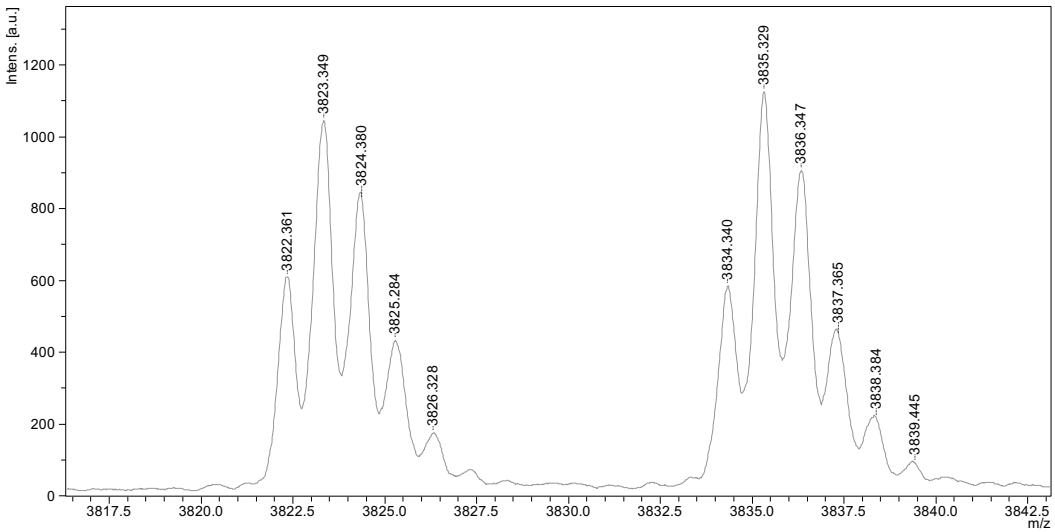


**Figure S21.** MALDI-TOF MS of **ON2** $m^4C$-12 mer (5′-GGACU$m^4C$CUGCAG-3′) $[M+H]^+$ = 3823.3 (calc. 3823.6).

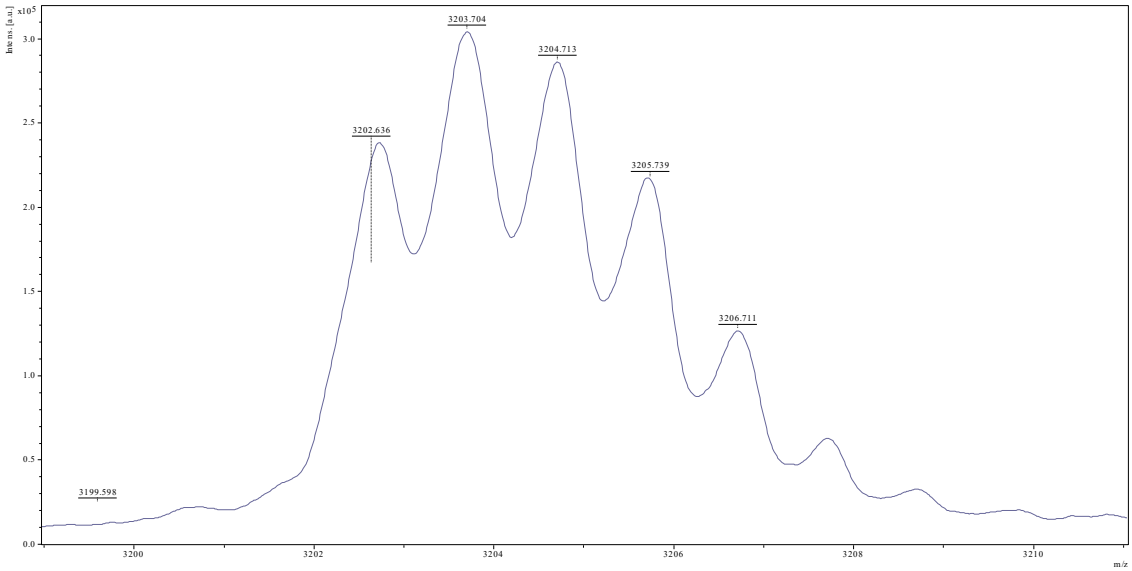


**Figure S22.** MALDI-TOF MS of **ON3** $m^4C$-10 mer (5′-CCGG$m^4C$GCCGG-3′) $[M+H]^+$ = 3203.7 (calc. 3203.5).

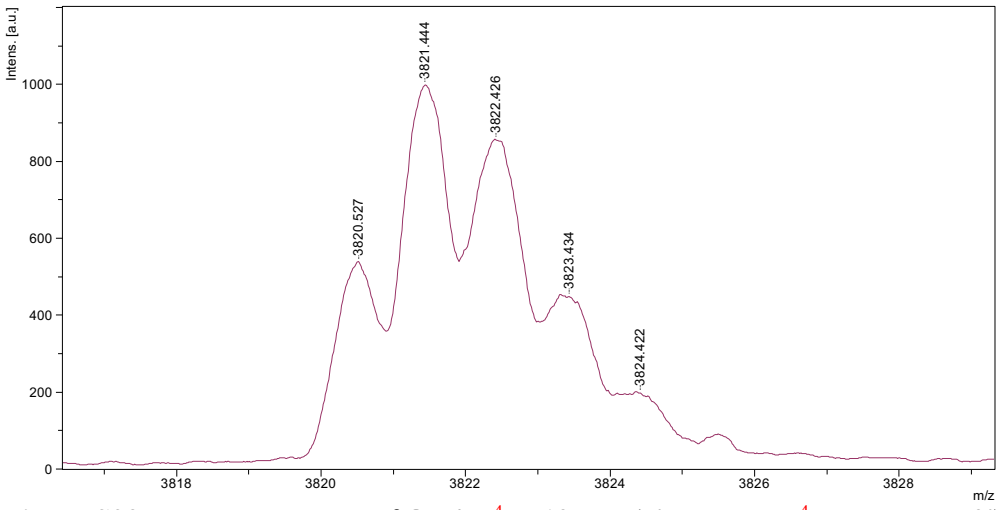


**Figure S23.** MALDI-TOF MS of **ON4** $m^4{}_2C$-12 mer (5′-AAUGC$m^4{}_2C$GCACUG-3′) $[M+H]^+$ = 3821.4 calc. 3821.6).

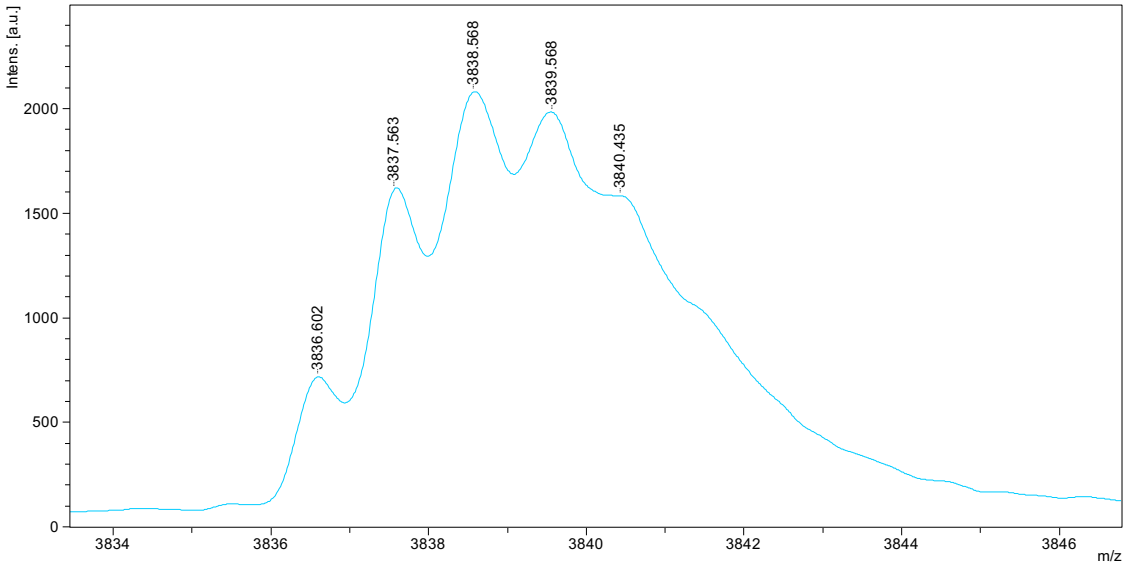


**Figure S24.** MALDI-TOF MS of **ON5** $m^4{}_2C$-12 mer (5′-GGACU$m^4{}_2C$CUGCAG-3′) $[M+H]^+$ = 3837.6 (calc. 3837.6).

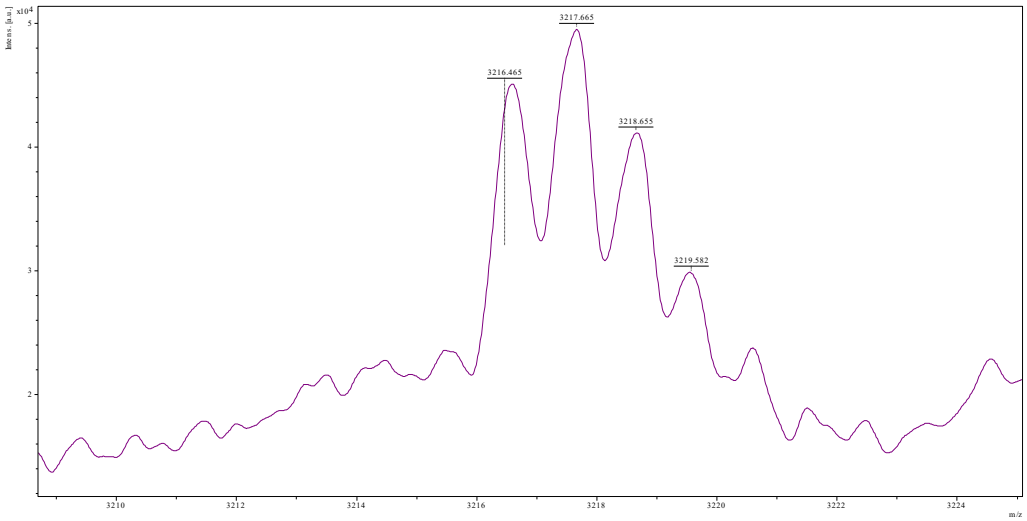


**Figure S25.** MALDI-TOF MS of **ON6** $m^4{}_2C$-10 mer (5′-CCGG$m^4{}_2C$GCCGG-3′) $[M+H]^+$ = 3217.7 calc.3217.5).

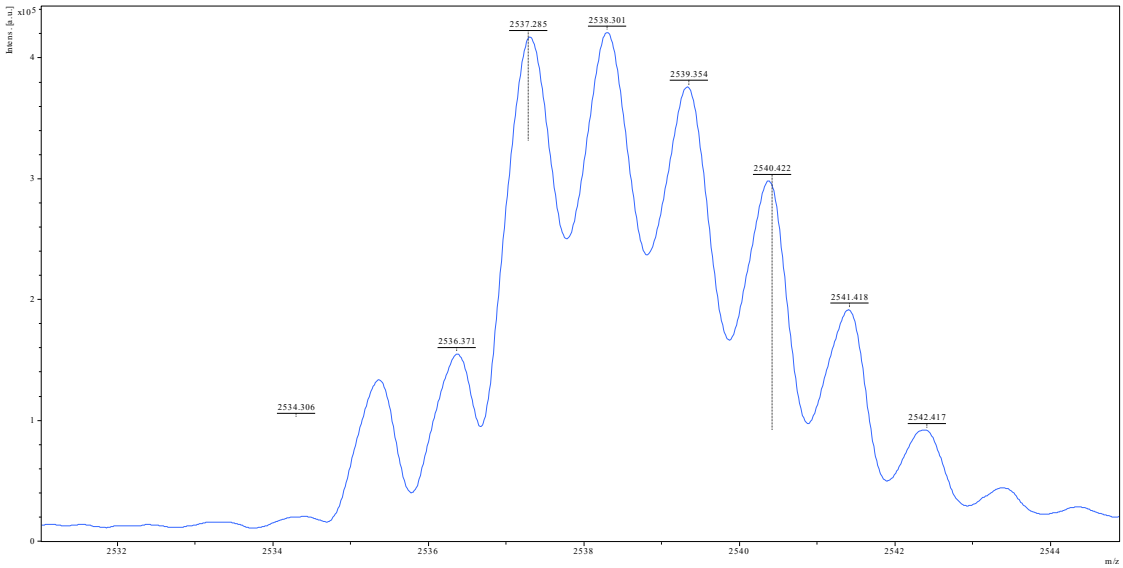


**Figure S26.** MALDI-TOF MS of **ON7** $m^4{}_2C$-8 mer (5′-U$m^4{}_2C$GUACGA-3′) $[M+H]^+$ = 2537.3 (calc.2537.4)

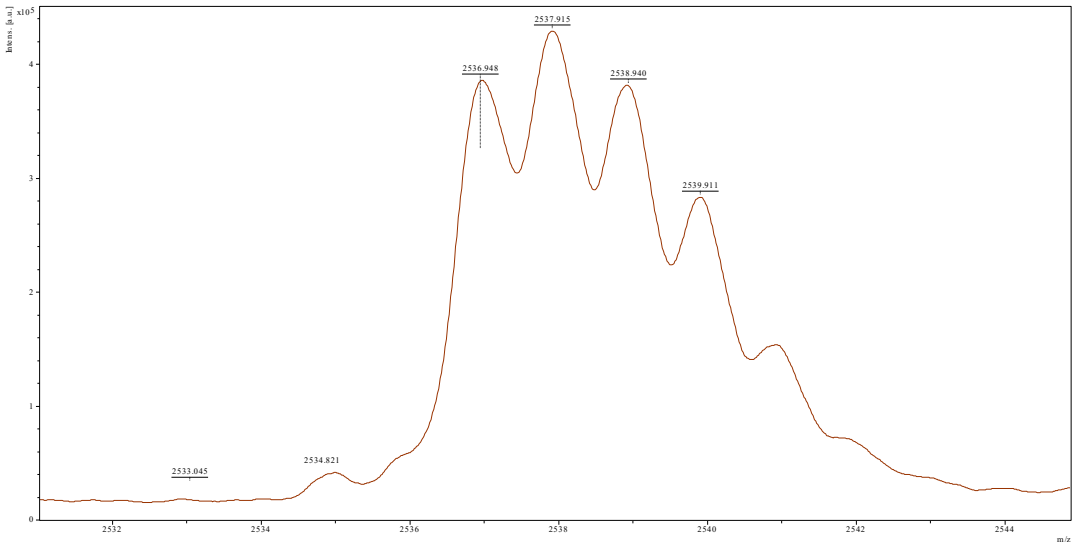


**Figure S27.** MALDI-TOF MS of **ON8** $m^4{}_2C$-8 mer (5′-GUA$m^4{}_2C$GUAC-3′) $[M+H]^+$ = 2537.9 (calc.2537.4)

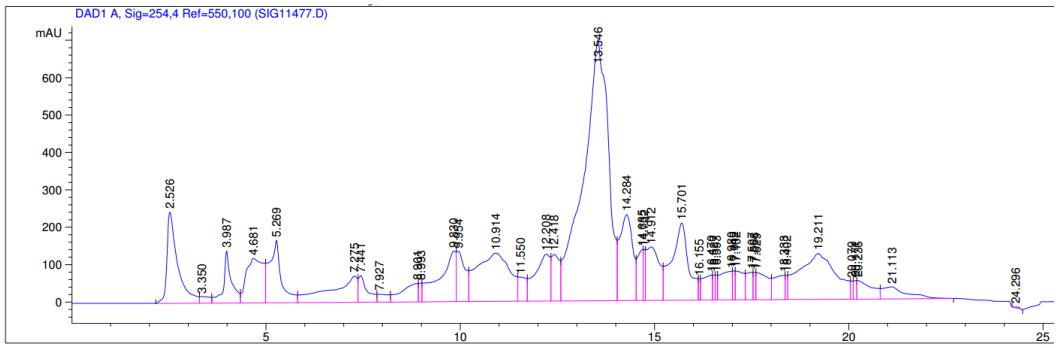


**Figure S28.** HPLC of **ON1** $m^4C$-12 mer (5′-AAUGC$m^4C$GCACUG-3′)

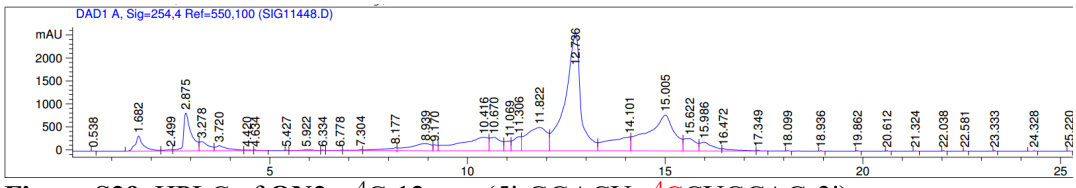


**Figure S29.** HPLC of **ON2** $m^4C$-12 mer (5′-GGACU$m^4C$CUGCAG-3′)

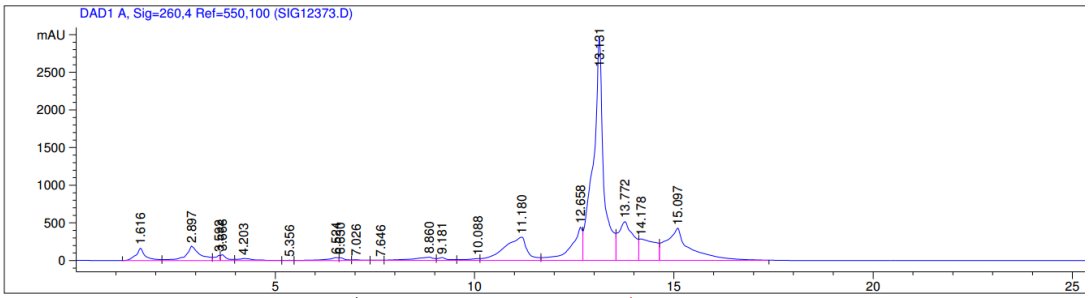


**Figure S30.** HPLC of **ON3** $m^4C$-10 mer (5′-CCGG$m^4C$GCCGG-3′)

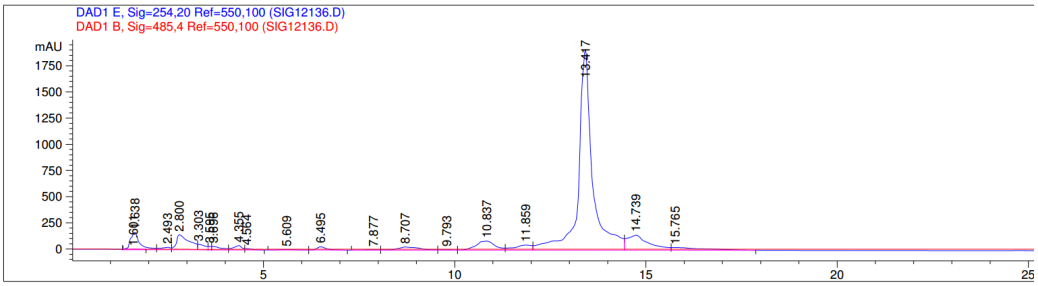


**Figure S31.** HPLC of **ON6** $m^4{}_2C$-10 mer (5′-CCGG$m^4{}_2C$GCCGG-3′)

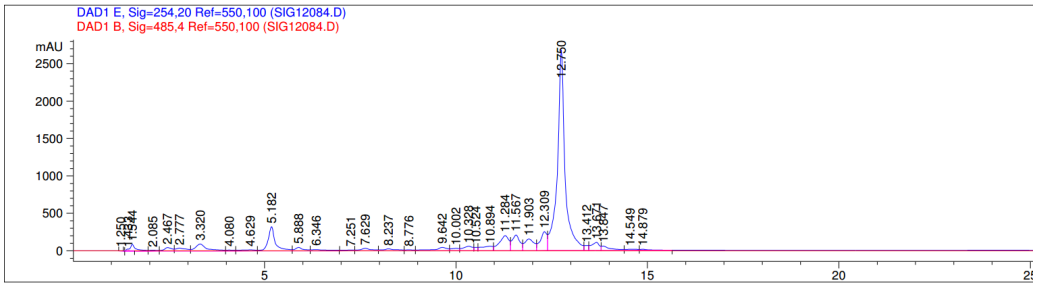


**Figure S32.** HPLC of **ON7** $m^4{}_2C$-8 mer (5′-U$m^4{}_2C$GUACGA-3′)

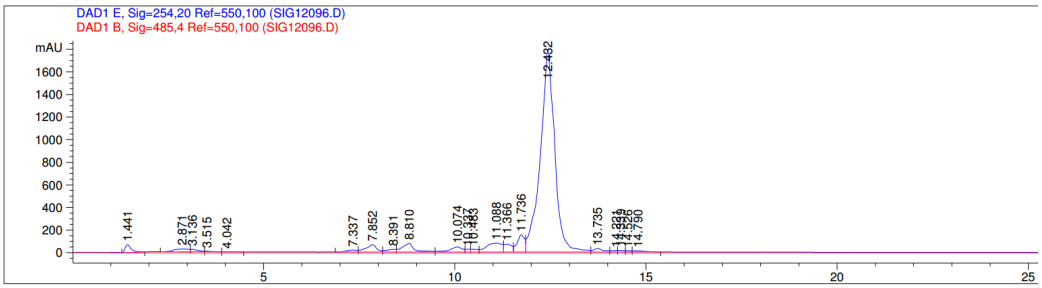


**Figure S33.** HPLC of **ON8** $m^4{}_2C$-8 mer (5′-GUA$m^4{}_2C$GUAC-3′)


**Figure S34.** 31-mer modified RNA templates after gel purification.

**UV-melting temperature ($T_m$) study**

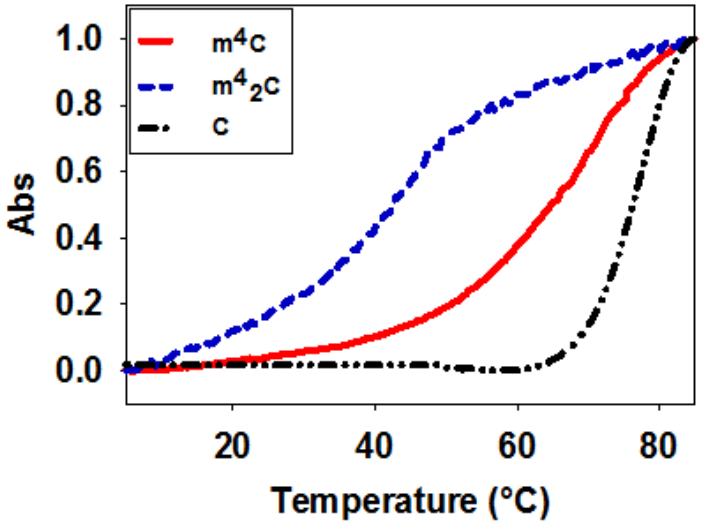


**Figure S35.** Normalized UV-melting curves of RNA duplexes (CCGGC*GCCGG)$_2$ with C* represents native C (black), m$^4$C (red) and m$^4{}_2$C (blue) residues.

**Table S2.** Melting temperatures of native, m$^4$C and m$^4{}_2$C modified 10-mer self-complementary RNA duplex (CCGGCGCCGG)$_2$.

| Entry | Sequences | Base pair | $T_m$ (°C)[a] | $\Delta T_m$ (°C)[b] | -$\Delta$G$^0$ (kcal/mol)[c] |
|---|---|---|---|---|---|
| 1 | (5′-CCGG**C**GCCGG-3′)$_2$ | C:G | 74.6 | | 18.3 |
| 2 | (5′-CCGG**m$^4$C**GCCGG-3′)$_2$ | m$^4$C:G | 66.9 | -7.7 | 11.6 |
| 3 | (5′-CCGG**m$^4{}_2$C**GCCGG-3′)$_2$ | m$^4{}_2$C:G | 42.4 | -32.2 | 8.5 |

[a] The $T_m$s were measured in sodium phosphate (10 mM, pH 7.0) buffer containing 100 mM NaCl, $T_m$ values reported are the averages of four measurements.
[b] $\Delta T_m$ values are relative to the duplexes with only Watson-Crick pairs.
[c] Obtained by non-linear curve fitting using Meltwin 3.5.

**X-ray crystal structure studies.**

**Table S3.** Data-collection and refinement statistics summary.

| Structure: | CCGG($m^4$C)GCCGG | CCGG($m^4{}_2$C)GCCGG | CCGG($m^4{}_2$C)GCCGG |
|---|---|---|---|
| **Data collection** | | | |
| Beamline | 22-ID | 22-ID | 22-ID |
| Wavelength (Å) | 1.00 | 1.00 | 1.00 |
| Temperature (K) | 100 | 100 | 100 |
| Oscillation range (°) | 0.25 | 0.25 | 0.25 |
| Space group | *C*2 | $P2_12_12_1$ | $R3_2$ |
| Unit cell parameters (Å,°) | a=94.7 b=30.4 c=58.5, β=106.5 | a=33.1 b=36.2 c=105.0 | a=b=42.6 c=165.9 |
| Resolution[1] (Å) | 28.05-1.93[1a] (2.00-1.93) | 28.00-1.65[1b] (1.69-1.65) | 35.98-1.81[1c] (1.99-1.81) |
| Reflections collected/unique | 27014/10372 (1437/518) | 49076/14279 (2198/709) | 31488/2771 (2674/231) |
| Completeness (%) | | | |
| Spherical | 83.9 (42.1) | 89.7 (65.9) | 49.2 (16.9) |
| Ellipsoidal | 88.9 (64.4) | 90.8 (80.4) | 85.8 (90.8) |
| Multiplicity | 2.6 (2.8) | 3.4 (3.1) | 11.4 (11.6) |
| $R_{merge}$ (%) | 4.8 (74.3) | 9.6 (67.7) | 7.1 (143.1) |
| $\langle I/\sigma(I)\rangle$ | 9.9 (1.7) | 6.7 (2.4) | 16.4 (1.8) |
| $CC_{1/2}$ (%) | 99.6 (64.5) | 99.3 (60.6) | 99.9 (77.3) |
| **Refinement** | | | |
| $R_{free}$ reflections | 504 | 692 | 149 |
| No. of atoms (non-H) | | | |
| RNA | 1070 | 856 | 428 |
| Ligands | 7 | 0 | 0 |
| Solvent | 139 | 187 | 3 |
| Molecules per asymmetric unit | Two duplexes and one single strand | Two duplexes | One duplex |
| $R_{work}/R_{free}$ (%) | 18.71/21.84 | 18.96/21.42 | 25.38/28.80 |
| Mean ADP[2] (Å$^2$) | 42.0 | 24.0 | 48.8 |
| RMSD from ideal geometry | | | |
| bond lengths (Å) | 0.004 | 0.006 | 0.006 |
| bond angles (°) | 0.84 | 1.16 | 1.12 |
| PDB code | 6WY2 | 6WY3 | 6Z18 |

*[1]Best anisotropic diffraction limit cut-off. [1a]Lowest cut-off diffraction limit is 2.29 Å. [1b]Lowest cut-off diffraction limit is 1.91 Å. [1c]Lowest cut-off diffraction limit is 2.75 Å. [2]ADP, atomic displacement parameter. Values in parentheses refer to the highest resolution shell.*

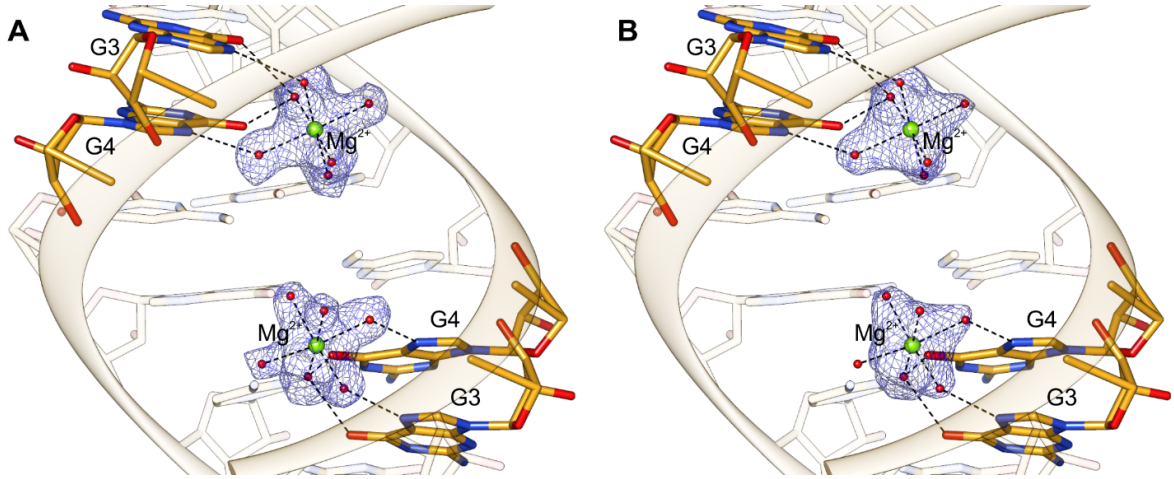


**Figure S36.** Coordination of the $Mg^{2+}$ ions in the CCGG($m^4$C)GCCGG structure. Side-to-side comparison of the final $2F_o$-$F_c$ electron density maps (blue mesh, displayed at 1 σ) after anisotropic cut-off in *STARANISO* (A) with the analogical map obtained for the spherically truncated data (B).

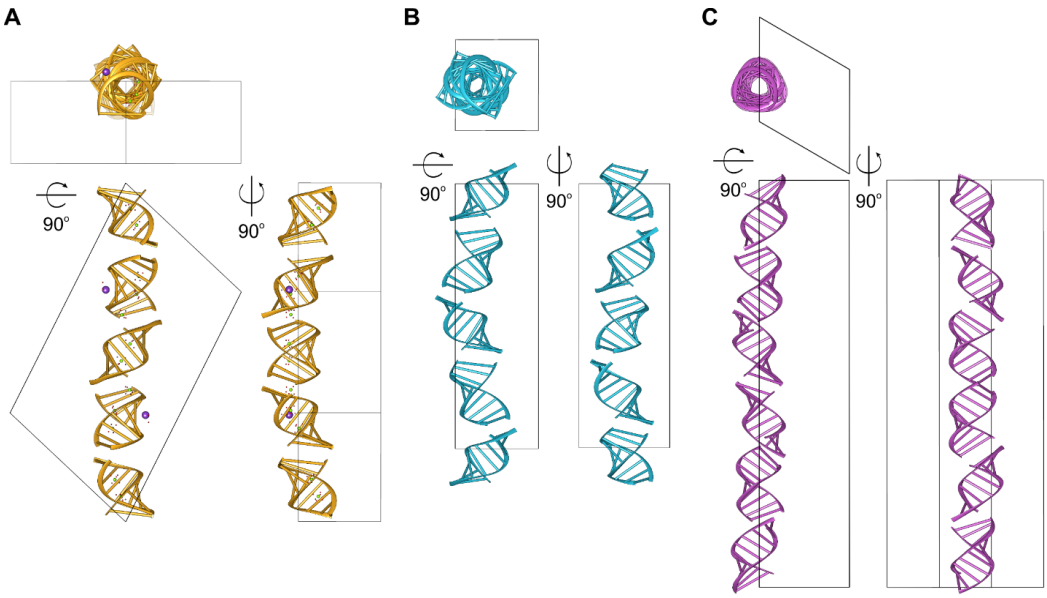


**Figure S37.** Crystal packing of the three solved structures. View of the helix in the unit cell created by the RNA 10-mers (A) CCGG($m^4$C)GCCGG, (B) CCGG($m^4{}_2$C)GCCGG in the $P2_12_12_1$ space group and (C) CCGG($m^4{}_2$C)GCCGG in the $R3_2$ space group.

**Time course gel images of primer extension reactions using HIV-1 RT with $m^4C$ and $m^4{}_2C$ modified templates.**

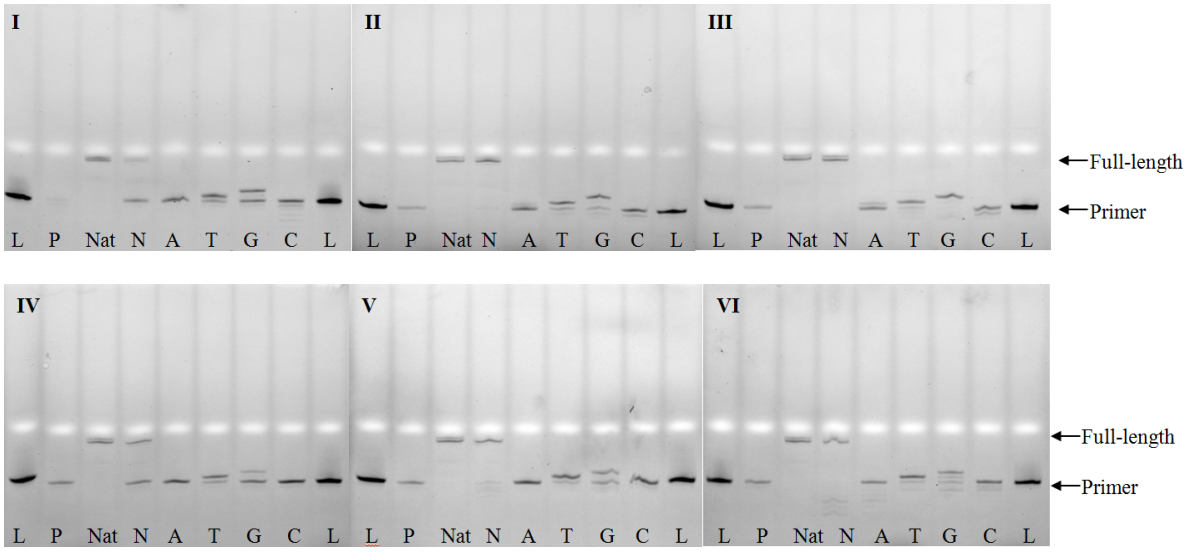


**Figure S38.** Time course fluorescent gel images of the standing-start primer extension reactions for HIV-1 RT using $m^4C$ (**I**: 0.5 h, **II**: 1 h and **III**: 2 h), $m^4{}_2C$ (**IV**: 0.5 h, **V**: 1 h and **VI**:1.5 h) containing RNA templates. Lanes: L, ladders; P, primer; Nat, natural template with all four dNTPs; A, T, G, and C, reactions in the presence of the respective dNTP; N, reactions in the presence of all four dNTPs.

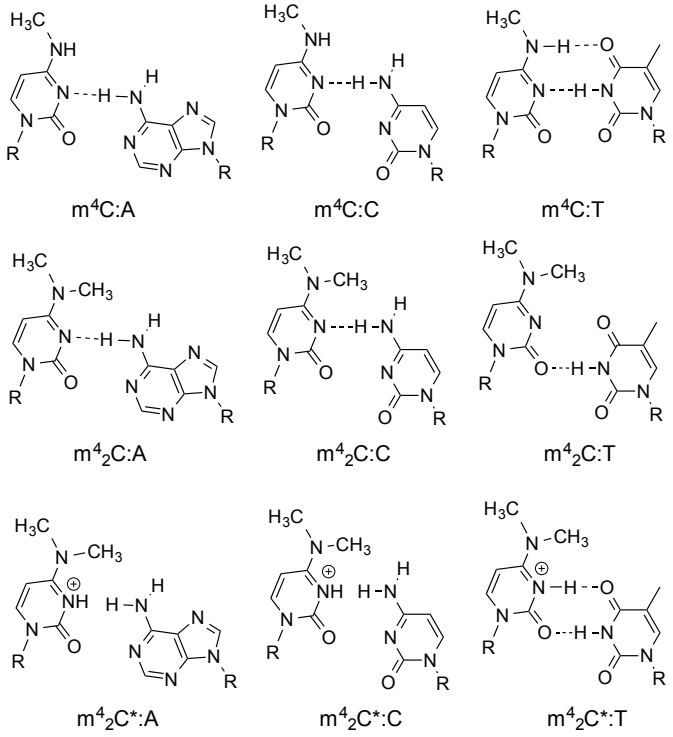


**Figure S39.** Base pairing patterns of $m^4C$ and $m^4{}_2C$ mispaired with A, C and T. $m^4{}_2C^*$ represents the protonated $C^+$ form.