Supporting Information

FRET Characterization of Complex Conformational Changes in a Large 16S Ribosomal RNA Fragment Site-Specifically Labeled Using Unnatural Base Pairs

Thomas Lavergne,[†#] Rajan Lamichhane,[‡#] Denis A. Malyshev,[†] Zhengtao Li,[†] Lingjun Li,[†] Edit Sperling,[‡] James R. Williamson,[‡] David P. Millar,[‡] Floyd Romesberg[†]*

[†]Department of Chemistry and [‡]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California, 92037

Table of Contents

Synthetic procedures and characterizations

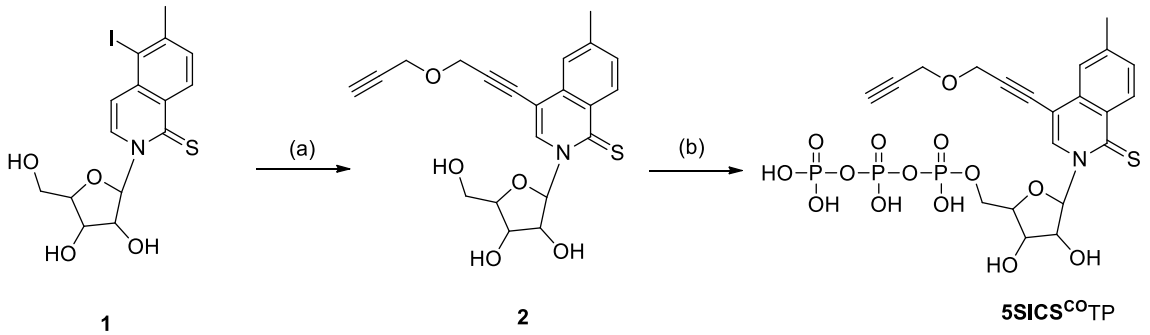
General.
All reactions were carried out in oven-dried glassware under inert atmosphere, and all solvents were dried over 4 Å molecular sieves with the exception of tetrahydrofuran, which was distilled from sodium metal. All other reagents were purchased from Aldrich or Fisher. ^{1}H, ^{13}C, and ^{31}P NMR spectra were recorded on Varian Mercury 300, Varian Inova-400, Bruker AMX-400, Bruker DRX-500 or Bruker DRX-600 spectrometers. High-resolution mass spectroscopic data were obtained on an ESI-TOF mass spectrometer (Agilent 6200 Series) at the TSRI Open Access Mass Spectrometry Lab; MALDI-TOF mass spectrometry (Applied Biosystems Voyager DE-PRO System 6008) was performed at the TSRI Center for Protein and Nucleic Acid Research.

T7 RNAP was obtained from Takara USA, and [α-^{32}P]GTP and [γ-^{32}P]GTP were obtained from MP Biomedicals. DNA ladders were obtained from Invitrogen.

General Procedure for Triphosphate Synthesis:
Proton sponge (1.3 equiv) and the free nucleoside derivative (1.0 equiv) were dissolved in dry trimethyl phosphate (40 equiv) and cooled to -15 °C under nitrogen atmosphere. Freshly distilled $POCl_3$ (1.3 equiv) was added dropwise and the resulting mixture was stirred at -10 °C for 2 h. Tributylamine (6.0 equiv) and a solution of tributylammonium pyrophosphate (5.0 eq.) in dimethylformamide (0.5 M) were added. Over 30 min, the reaction was allowed to warm slowly to 0 °C and then was quenched by addition of 0.5 M aqueous $Et_3NH_2CO_3$ (TEAB) pH 7.5 (2 vol-equiv.). The mixture was diluted two-fold with H_2O and the

product was isolated on a DEAE Sephadex column (GE Healthcare) with an elution gradient of 0 to 1.2 M TEAB, evaporated, and co-distilled with H_2O (3×). Additional purification by reverse-phase (C18) HPLC (0 - 35% CH_3CN in 0.1 M TEAB, pH 7.5) was performed.

Synthetic schemes and procedures:



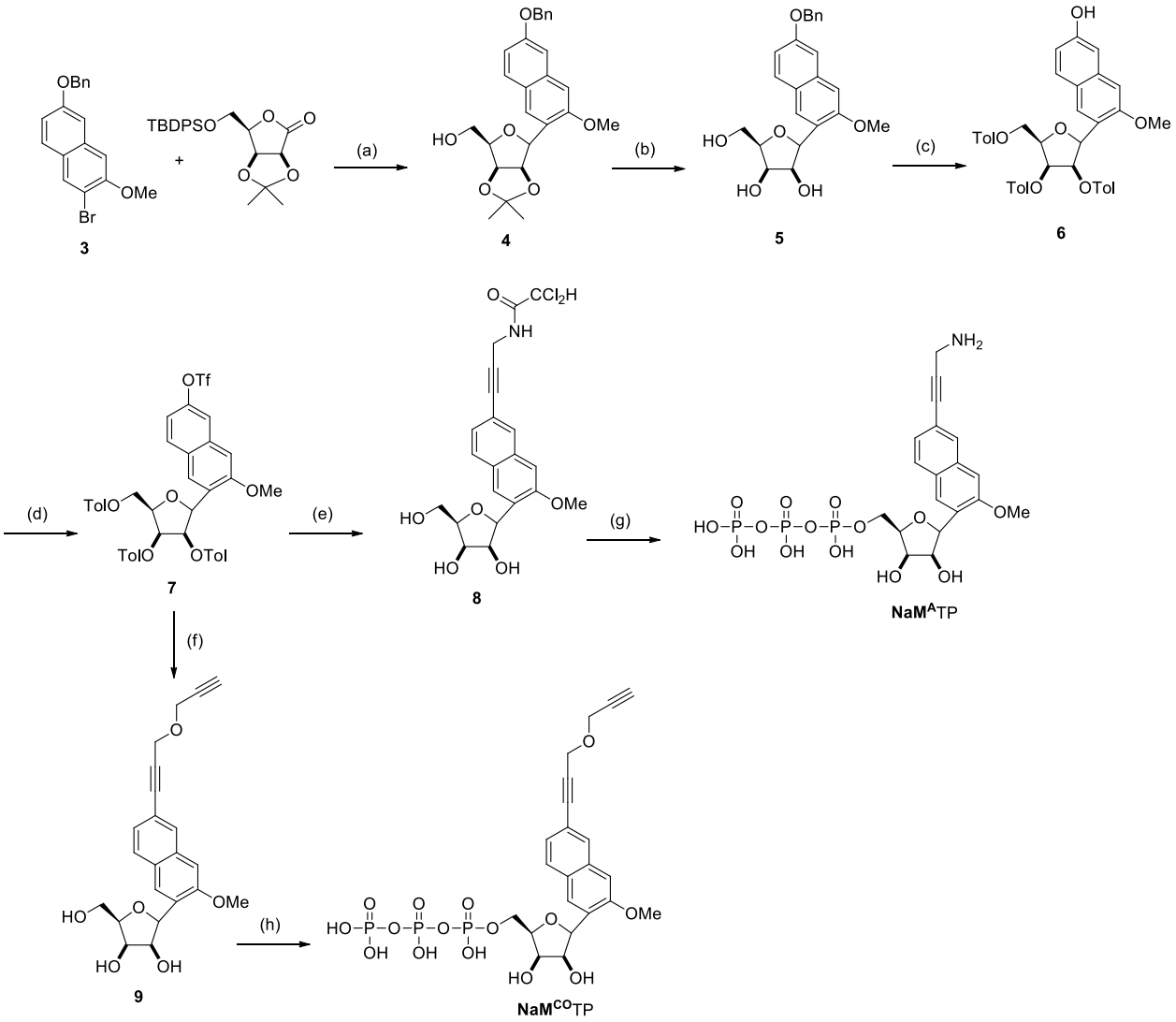
1 2 **5SICS^{CO}TP**

Scheme S1: (a) Propargyl ether, $(PPh_3)_4Pd$, CuI, Et_3N, DMF, rt, 32%; (b) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 22%.

Compound 1: Compound **1** was synthesized as described previously.[1]

Compound 2: To a solution of **1** (52 mg, 0.120 mmol) in DMF (16 mL) under nitrogen atmosphere was added $(PPh_3)_4Pd$ (13.6 mg, 0.012 mmol), CuI (4.6 mg, 0.088 mmol) and Et_3N (40 µl, 0.24 mmol). The reaction mixture was degassed and a solution of propargyl ether (0.220 mmol) in DMF (4 mL) was added. The reaction mixture was stirred overnight at room temperature and monitored by TLC. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous $NaHCO_3$, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography ($MeOH/CH_2Cl_2$) to afford compound **2** as yellow solid (15.2 mg, 0,038 mmol, 32%). ^{1}H NMR (500 MHz, CD_3OD) δ 8.83 (d, J = 8.5 Hz, 1 H), 8.75 (s, 1 H), 7.72 (s, 1H), 7.38 (d, J = 8.5 Hz, 1H), 6.94 (s, 1H), 4.59 (s, 2 H), 4.37 (d, J = 0.5 Hz, 2H), 4.26-4.16 (m, 3H), 4.19 (d, J = 12.5 Hz, 1 H), 3.88 (d, J = 12.5 Hz, 1 H). 2.95 (t, J = 1.5 Hz, 1 H), 2.49 (s, 3 H). ^{13}C NMR (125 MHz, CD_3OD) δ 182.7, 144.5, 133.4, 131.8, 131.7, 130.3, 124.5, 105.5, 96.0, 89.8, 84.3, 80.9, 78.9, 76.0, 75.6, 68.2, 59.5, 57.0, 56.5, 20.8. HRMS (ESI+) *m/z* calcd for $C_{21}H_{22}NO_5S$ ($M+H^+$) 400.1213, Found 400.1213

5SCIS^{CO}TP: **5SCIS^{CO}TP** (4.2 mg, 0.0066 mmol, 22%) was synthesized using the general procedure for triphosphate synthesis starting from **2** (12 mg, 0.03 mmol). ^{31}P NMR (162 MHz, D_2O) δ -10.02 (d, J = 19.8 Hz, γ-P), -11.58 (d, J = 20 Hz, α-P), -23.04 (t, J = 20 Hz, β-P). MS (MALDI-TOF-, matrix: 9-aminoacridine) (*m/z*): $[M-H]^-$ calcd for $C_{21}H_{23}NO_{14}P_3S$ 638.0, found, 638.1.



Scheme S2: (a) i/ *t*buLi, -78 °C, 10 min ii/ lactone, -78 °C, 3 h iii/ Et_3SiH/BF_3-Et_2O iv/ TBAF,THF. 24% (four steps) (b) 70% AcOH, 100 °C, 30 min, 88% (c) i/ 4-methylbenzoic anhydride, DMAP, pyridine ii/ Pd/C, H_2, EtOAc. 88% (two steps) (d) Tf_2O, DMAP, pyridine, 92% (e) i/ $PdCl_2(PPh_3)_2$, CuI, nBu_4NI, Et_3N, 40 °C, propargyl ether ii/ NaOMe, MeOH. 85% (two steps) (f) $PdCl_2(PPh_3)_2$, CuI, nBu_4NI, Et_3N, 40 °C, 2,2-dichloro-N-(prop-2-yn-1-yl)acetamide ii/ NaOMe, MeOH. 80% (two steps) (g) i/ Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C ii/ NaOH 0,1M, rt, 2 h. 30% (two steps) (h) Proton sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, -20 °C, 50%.

Compound 3: 6-(benzyloxy)-2-bromo-3-methoxynaphthalene (compound **3**) was synthesized as described previously.[2]

Compound 4: To a solution of compound **3** (342 mg, 1.0 mmol) in 4 mL of THF was added *t*BuLi (1.6M, 0.9 mmol) at -78 °C. The resulting mixture was stirred for 10 min at -78 °C and the silyl lactone (217 mg, 0.50 mmol) in 2 mL THF was added slowly. After addition, the reaction mixture was stirred at -78 °C for another 3 h. The reaction was monitored by TLC and quenched by addition of saturated NH_4Cl solution, extracted with ethyl acetate (2 × 20 mL) and washed with brine, dried over anhydrous Na_2SO_4, filtered and evaporated. The residue was isolated by filtration over silica gel (EtOAc/hexane =1/3) and subjected to the subsequent step without further purification. The residue (160 mg, 0.23 mmol) was dissolved in CH_3CN (3 mL) and the solution was cooled to -40 °C. Et_3SiH was added (74 µL, 0.46

mmol) followed by the addition of $BF_3 \cdot Et_2O$ (35 µL, 0.28 mmol). The resulting mixture was stirred at -40 °C for 1 h then warmed up to room temperature and quenched by addition of saturated K_2CO_3 solution. The product was extracted with ethyl acetate (2 × 20 mL) and the organic layer was washed with brine, dried over anhydrous Na_2SO_4, filtered and evaporated. The residue was isolated by filtration over silica gel (EtOAc/hexane =1/4) and subjected to the subsequent step without further purification. To the residue (128 mg, 0.19 mmol) in THF (2 mL) was added TBAF (1 M, 768 µL) and the mixture was stirred for 2 h. The reaction was monitored by TLC and quenched by addition of saturated NH_4Cl solution, extracted with ethyl acetate (2 × 10 mL) and washed with brine, dried over anhydrous Na_2SO_4, filtered and evaporated. The residue was purified by a flash chromatography on silica gel (EtOAc/hexane =1/2) and the desired compound **4** was obtained as white foam after evaporation of the solvent (53.1 mg, 0.122 mmol, 24%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.92 (s, 1H), 7.71 (d, *J* = 8.9 Hz, 1H), 7.49 (d, *J* = 7.0 Hz, 2H), 7.39 (m, 4H), 7.14 (d, *J* = 2.4 Hz, 1H), 7.08 (dd, *J* = 8.8, 2.5 Hz, 1H), 7.02 (s, 1H), 5.36 (d, *J* = 4.0 Hz, 1H), 5.17 (s, 2H), 5.04 (dd, *J* = 6.0, 4.0 Hz, 1H), 4.74 (dd, *J* = 6.0, 1.0 Hz, 1H), 4.42 (dd, *J* = 7.4, 4.8 Hz, 1H), 3.93 (s, 3H), 3.83 – 3.68 (m, 3H), 1.35 (s, 3H), 1.26 (s, 3H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 157.97, 155.77, 137.90, 135.93, 130.32, 129.45, 128.80, 128.36, 127.77, 124.96, 124.80, 117.14, 113.40, 107.33, 104.96, 84.72, 83.40, 82.36, 78.84, 70.82, 62.51, 56.28, 30.56, 30.11, 27.12, 26.01. HRMS (ESI+) *m/z*: calcd for $C_{26}H_{28}O_6$ [M+H]+ 437.1959, found 437.1956.

Compound 5: To compound **4** (150 mg, 0.344 mmol) was added a solution of 70% CH_3COOH in H_2O (4 mL), the resulting mixture was heated to 100 °C for 30 min. The reaction was monitored by TLC and compound **5** was extracted with ethyl acetate (2 × 10 mL). The organic layer was washed with brine then saturated $NaHCO_3$, dried over anhydrous Na_2SO_4, filtered and evaporated. The residue was purified by a flash chromatography on silica gel (EtOAc/hexane =2/1) and the desired compound was obtained as white foam after evaporation of the solvent (120 mg, 0.302 mmol, 88%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.88 (s, 1H), 7.69 (d, *J* = 8.9 Hz, 1H), 7.48 (m, 2H), 7.39 (m, 4H), 7.13 (d, *J* = 2.4 Hz, 1H), 7.09 (dd, *J* = 8.8, 2.5 Hz, 1H), 6.99 (s, 1H), 5.52 (dd, *J* = 3.2, 0.9 Hz, 1H), 5.15 (s, 2H), 4.45 (t, *J* = 3.9 Hz, 1H), 4.35 (dd, *J* = 8.1, 4.5 Hz, 1H), 4.06 (dd, *J* = 7.7, 3.8 Hz, 1H), 3.97 (dd, *J* = 12.0, 3.0 Hz, 1H), 3.87 (s, 2H), 3.76 (dd, *J* = 12.0, 4.1 Hz, 1H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 158.31, 155.76, 137.73, 136.20, 130.33, 129.48, 128.87, 128.35, 127.77, 124.92, 124.64, 117.61, 107.28, 105.40, 83.04, 79.90, 74.26, 73.01, 70.86, 63.41, 56.30, 54.59, 32.59, 30.10. HRMS (ESI^+) *m/z*: calcd for $C_{23}H_{24}O_6$ $[M+H]^+$ 397.1646, found 397.1650.

Compound 6: To a solution of compound **5** (40 mg, 0.1 mmol) in pyridine (2 mL) was added DMAP (3.66 mg, 0.03 mmol). 4-methylbenzoic anhydride (152.6 mg, 0.6 mmol) in DCM (2 mL) was then added dropwise at 0 °C and the resulting mixture was stirred overnight at room temperature. The reaction was monitored by TLC and quenched by addition of H_2O, extracted with ethyl acetate (2 × 10 mL) and washed with brine followed by saturated NaHCO3, dried over anhydrous Na2SO4, filtered and evaporated. The crude product was dried under vacuum and was used without further purification for the following step. To the crude residue in EtOAc (4 mL) was added 10% palladium on carbon (20 mg) under the hydrogen atmosphere, the reaction was stirred at room temperature for 20 h. The mixture was passed through a pad of silica, the solvent was removed and the remaining crude mixture was subjected to silica gel column chromatography (EtOAc/hexane =2/1) to afford compound **6** as a white foam (58.0 mg, 0.088 mmol, 88%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.09 – 7.93 (m, 4H), 7.69-7.66 (m, 3H), 7.41 (d, *J* = 8.2 Hz, 2H), 7.19 (d, *J* = 8.0 Hz, 2H), 7.03 (d, *J* = 8.0 Hz, 2H), 6.99 – 6.92 (m, 4H), 6.77 (s, 1H), 6.26 (t, *J* = 4.2 Hz, 1H), 5.96 (dd, *J* = 7.6, 4.6 Hz, 1H), 5.86-5.85 (m, 1H), 4.85 (dt, *J* = 7.8, 4.2 Hz, 1H), 4.79 (dd, *J* = 11.8, 3.8 Hz, 1H), 4.65 (dd, *J* = 11.8, 4.5 Hz, 1H), 3.79 – 3.65 (s, 3H), 2.38 (s, 3H), 2.32 (s, 3H), 2.26 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.49, 166.64, 165.75, 165.41, 155.22, 154.41, 144.04, 143.96, 143.69, 135.55, 129.93, 129.89, 129.65, 129.55, 129.29, 129.22, 129.05, 128.96, 127.09, 126.90, 126.64, 126.43, 123.90, 123.82, 115.52, 108.61, 103.62, 77.76, 77.72, 77.36, 73.87, 72.95, 64.51, 60.61, 55.33,

29.32, 21.78, 21.73, 21.64, 21.17. HRMS (ESI^+) *m/z*: calcd for $C_{40}H_{36}O_9$ $[M+H]^+$ 661.2432, found 661.2435.

Compound 7: To a solution of compound **6** (47 mg, 0.078 mmol) in pyridine (2 mL) was added DMAP (2.8 mg, 0.023 mmol). Triflate anhydride in DCM (1M, 235 µL) was then added dropwise at 0 °C and the resulting mixture was stirred overnight at room temperature. The reaction was monitored by TLC and quenched by addition of H_2O, extracted with ethyl acetate (2 × 10 mL) and washed with brine followed by saturated $NaHCO_3$, dried over anhydrous Na_2SO_4, filtered and evaporated. The residue was purified by a flash chromatography on silica gel (EtOAc/hexane =1/2) and the desired compound **7** was obtained as white foam after evaporation of the solvent (56 mg, 0.072 mmol, 92%). ^1H NMR (500 MHz, CDCl3) δ 8.27 (s, 1H), 8.06 (d, *J* = 8.2 Hz, 2H), 7.95 (d, *J* = 9.0 Hz, 1H), 7.74 (d, *J* = 8.2 Hz, 2H), 7.66 (d, *J* = 2.3 Hz, 1H), 7.44 (d, *J* = 8.2 Hz, 2H), 7.33 – 7.27 (m, 5H), 7.11 (d, *J* = 8.2 Hz, 2H), 7.08 (s, 1H), 7.02 (d, *J* = 8.1 Hz, 2H), 6.41 (t, *J* = 4.4 Hz, 1H), 6.04 (dd, *J* = 7.5, 4.7 Hz, 1H), 5.96 (d, *J* = 4.0 Hz, 1H), 4.95 (dt, *J* = 7.8, 4.2 Hz, 1H), 4.89 (dd, *J* = 11.9, 3.8 Hz, 1H), 4.74 (dd, *J* = 11.9, 4.5 Hz, 1H), 3.91 (s, 3H), 2.49 (s, 3H), 2.41 (s, 3H), 2.36 (s, 3H). ^{13}C NMR (151 MHz, CDCl3) δ 172.01, 167.21, 166.35, 165.78, 156.71, 148.43, 144.88, 144.75, 144.63, 135.07, 131.13, 130.65, 130.59, 130.16, 129.98, 129.80, 129.70, 129.37, 128.32, 127.79, 127.47, 127.37, 127.06, 118.55, 118.01, 105.64, 78.69, 78.43, 74.39, 73.50, 65.05, 61.25, 56.41, 22.55, 22.46, 22.37, 21.91, 15.05. HRMS (ESI^+) *m/z*: calcd for $C_{41}H_{35}F_3O_{11}S$ $[M+H]^+$ 793.1925, found 793.1925.

Compound 8: To a solution of compound **7** (16 mg, 0.02 mmol) in CH_3CN (1 mL) was added $PdCl_2(PPh_3)_2$ (6.0 mg, 0.006 mmol), nBu_4NI (14.8 mg, 0.04 mmol), dichloro-N-(prop-2-yn-1-yl)acetamide (10 mg, 0.06 mmol) and Et_3N (400 µL). The resulting mixture was degassed with argon several times and was stirred for 10 min. CuI (1.0 mg, 0.002 mmol) was then added and the reaction mixture was degassed with argon several times, warmed to 40 °C and stirred overnight. The reaction mixture was passed through a pad of silica and the washed with ethyl acetate and then the solvent was removed under vacuum to give a white foam crude residue. To a solution of this residue in a mixture of MeOH (2 mL) and DCM (1 mL) was added NaOMe (6 M, 40 µL). The reaction was monitored by TLC and quenched by addition of saturated NH_4Cl solution, extracted with ethyl acetate (2 × 10 mL) and washed with brine, dried over anhydrous Na_2SO_4, filtered and evaporated. The residue was purified by a flash chromatography on silica gel (EtOAc/hexane =3/1) and the desired compound **8** was obtained as white foam after evaporation of the solvent (7.5 mg, 0.017 mmol, 85% for two steps) ^1H NMR (400 MHz, MeOD) δ 7.95 (s, 1H), 7.86 (s, 1H), 7.72 (d, *J* = 8.4 Hz, 1H), 7.29 (dd, *J* = 8.4, 1.6 Hz, 1H), 7.17 (s, 1H), 6.30 (s, 1H), 5.38 (dd, *J* = 3.1, 1.0 Hz, 1H), 4.43 (t, *J* = 3.8 Hz, 1H), 4.32 (dd, *J* = 8.6, 4.5 Hz, 1H), 4.31(s,2H), 4.07 (dq, *J* = 8.5, 2.5 Hz, 1H), 3.97 – 3.91 (m, 4H), 3.76 – 3.72 (dd, *J* = 12.0, 4.7 Hz, 1H). ^{13}C NMR (151 MHz, MeOD-d4) δ 164.45, 156.60, 134.36, 131.08, 130.79, 128.94, 128.67, 127.64, 126.70, 120.51, 105.09, 85.54, 84.05, 83.38, 79.61, 73.74, 73.24, 67.46, 63.08, 55.86, 55.43. HRMS (ESI^+) *m/z*: calcd for $C_{21}H_{21}Cl_2NO_6$ $[M+H]^+$ 454.0819, found 454.0819.

Compound 9: To a solution of compound **7** (16 mg, 0.02 mmol) in CH_3CN (1 mL) was added $PdCl_2(PPh_3)_2$ (6.0 mg, 0.006 mmol), nBu_4NI (14.8 mg, 0.04 mmol), propargyl ether (200 µL, 1.94 mmol) and Et_3N (400 µL). The resulting mixture was degassed with argon several times and was stirred for 10 min. CuI (1.0 mg, 0.002 mmol) was then added and the reaction mixture was degassed with argon several times, warmed to 40 °C and stirred overnight. The reaction mixture was passed through a pad of silica and the washed with ethyl acetate and then the solvent was removed under vacuum to give a white foam crude residue. To a solution of this residue in a mixture of MeOH (2 mL) and DCM (1mL) was added NaOMe (6M, 40 µL). The reaction was monitored by TLC and quenched by addition of saturated NH_4Cl solution, extracted with ethyl acetate (2 × 10 mL) and washed with brine, dried over anhydrous Na_2SO_4, filtered and evaporated. The residue was purified by a flash chromatography on silica gel (EtOAc/hexane =3/1) and the desired compound **9** was obtained as white foam after evaporation of the solvent (6.1 mg, 0.016 mmol, 80% for two steps). ^1H NMR (400 MHz, acetone) δ 7.96 (s, 1H), 7.93 (s, 1H), 7.81 (d, *J* =

8.4 Hz, 1H), 7.35 (dd, *J* = 8.4, 1.6 Hz, 1H), 7.25 (s, 1H), 5.35 (dd, *J* = 3.1, 1.0 Hz, 1H), 4.52 (s, 2H), 4.47 (t, *J* = 3.8 Hz, 1H), 4.39 (m, 1H), 4.35 (d, *J* = 2.4 Hz, 2H), 4.05 (dt, *J* = 7.7, 3.5 Hz, 1H), 3.94 (s, 3H), 3.90 (m,1H), 3.74 – 3.66 (dd, *J* = 11.7, 3.8 Hz, 1H), 3.10 – 3.01 (t, *J* = 2.4 Hz, 1H). ^{13}C NMR (151MHz, Acetone) δ 156.63, 134.36, 131.22, 130.86, 129.03, 128.69, 127.66, 126.67, 120.33, 105.09, 87.69, 85.29, 83.48, 80.08, 79.61, 76.46, 73.71, 73.23, 63.06, 57.59, 56.87, 55.83. HRMS (ESI$^+$) *m/z*: calcd for $C_{22}H_{22}O_6$ $[M+Na]^+$ 405.1309, found 405.1307.

NaM^{A}TP: Compound **8** (12 mg, 0.0264 mmol) was subjected to the general procedure for triphosphate synthesis. The resulting compound was treated with a solution of NaOH 0.1 M (1 mL) for 2 h and purified by reverse-phase (C18) HPLC (0 - 35% CH3CN in 0.1 M TEAB, pH 7.5) providing **NaM^A**TP (4,6 mg, 0.0079 mmol, 30%). ^{31}P NMR (162 MHz, D_2O) δ -9.68 (d, *J* = 20,1 Hz, γ-P), -10.68 (d, *J* = 20 Hz, α-P), -22.76 (t, *J* = 19,9 Hz, β-P). MS (MALDI-TOF-, matrix: 9-aminoacridine) (*m/z*): $[M-H]^-$ calcd for $C_{19}H_{23}NO_{14}P_3$ 582.3, found, 582.5.

NaM^{CO}TP: **NaM^{CO}**TP (9.3 mg, 0.015 mmol, 50%) was synthesized using the general procedure for triphosphate synthesis starting from **9** (11.5 mg, 0.03 mmol). ^{31}P NMR (162 MHz, D_2O) δ -9.20 (d, *J* = 19,9 Hz, γ-P), -10.65 (d, *J* = 19,7 Hz, α-P), -22.67 (t, *J* = 20 Hz, β-P). MS (MALDI-TOF-, matrix: 9-aminoacridine) (*m/z*): $[M-H]^-$ calcd for $C_{22}H_{24}O_{15}P_3$ 621.3, found, 621.3.

NMR spectra and MALDI-TOF analysis

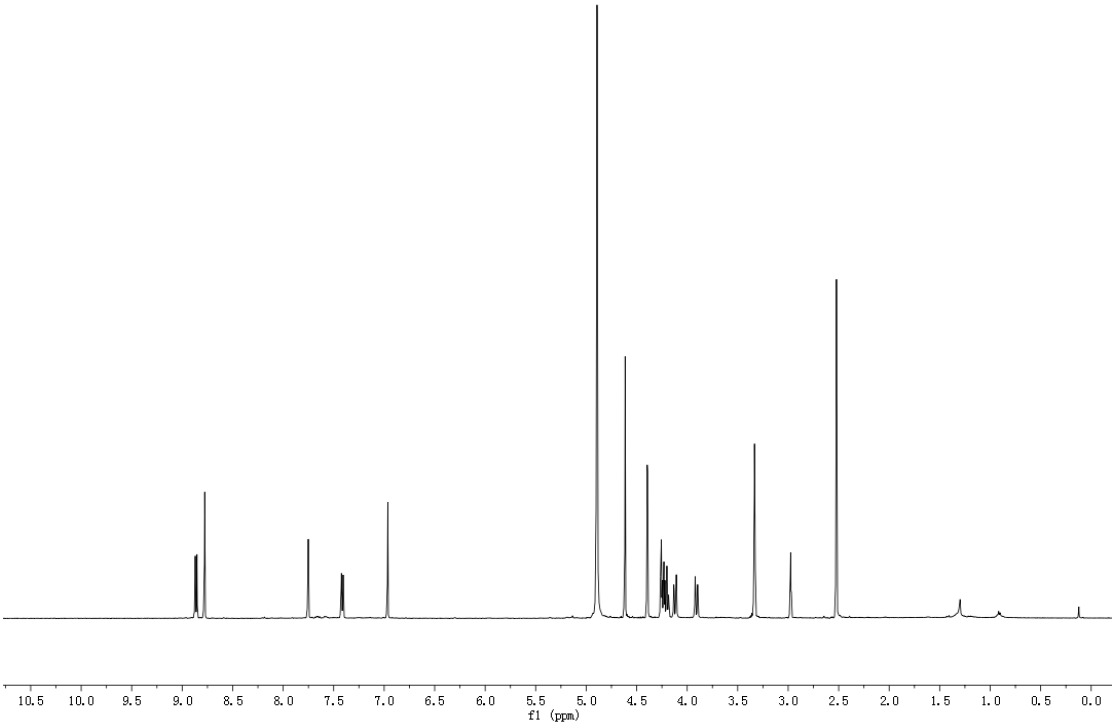


Figure S1: 500 MHz ^{1}H NMR spectra (CD_3OD) of **2**

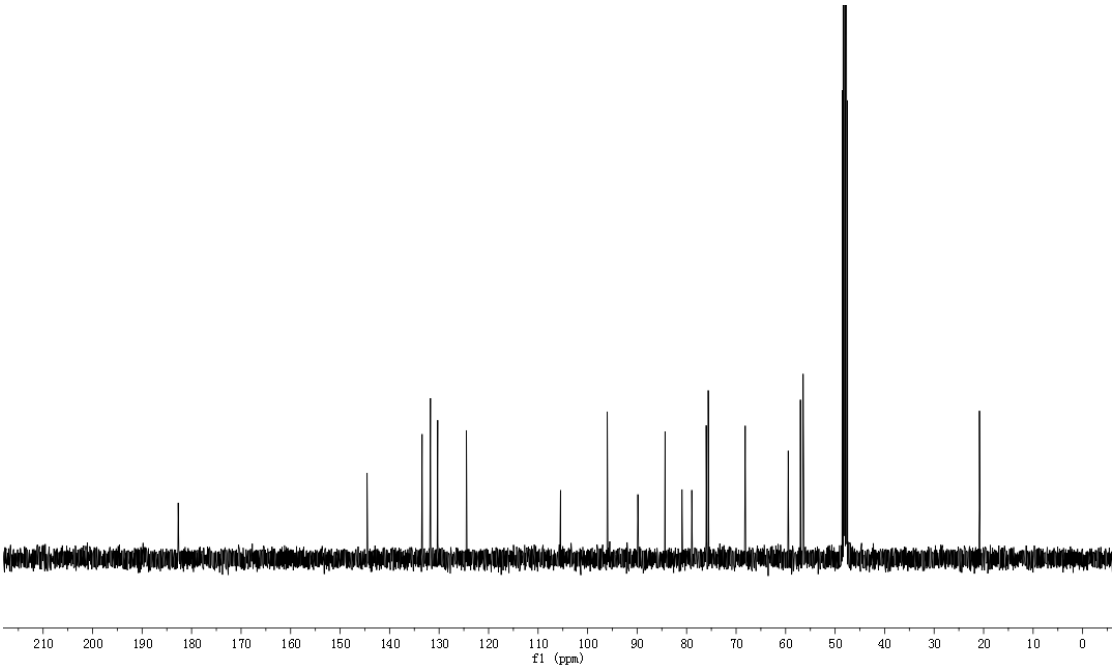


Figure S2: 125 MHz ^{13}C NMR spectra (CD_3OD) of **2**

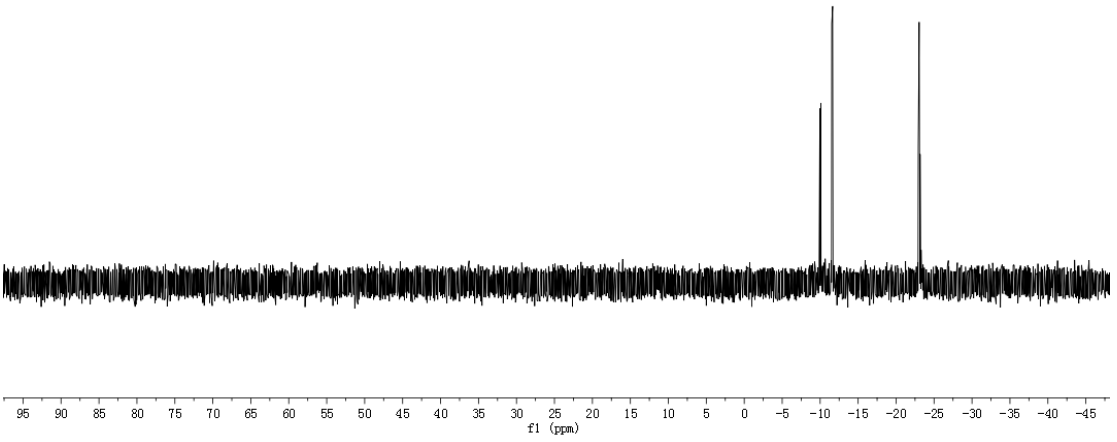


Figure S3: 162 MHz ^{31}P NMR spectra (D_2O) of **5SICS^{CO}TP**

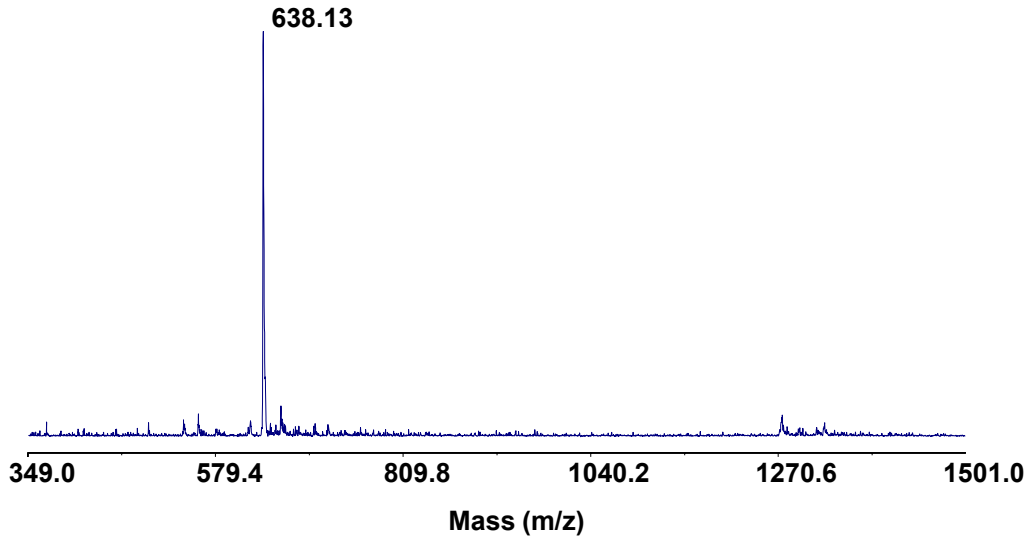


Figure S4: MALDI-TOF MS analysis of **5SICS**^{CO}TP

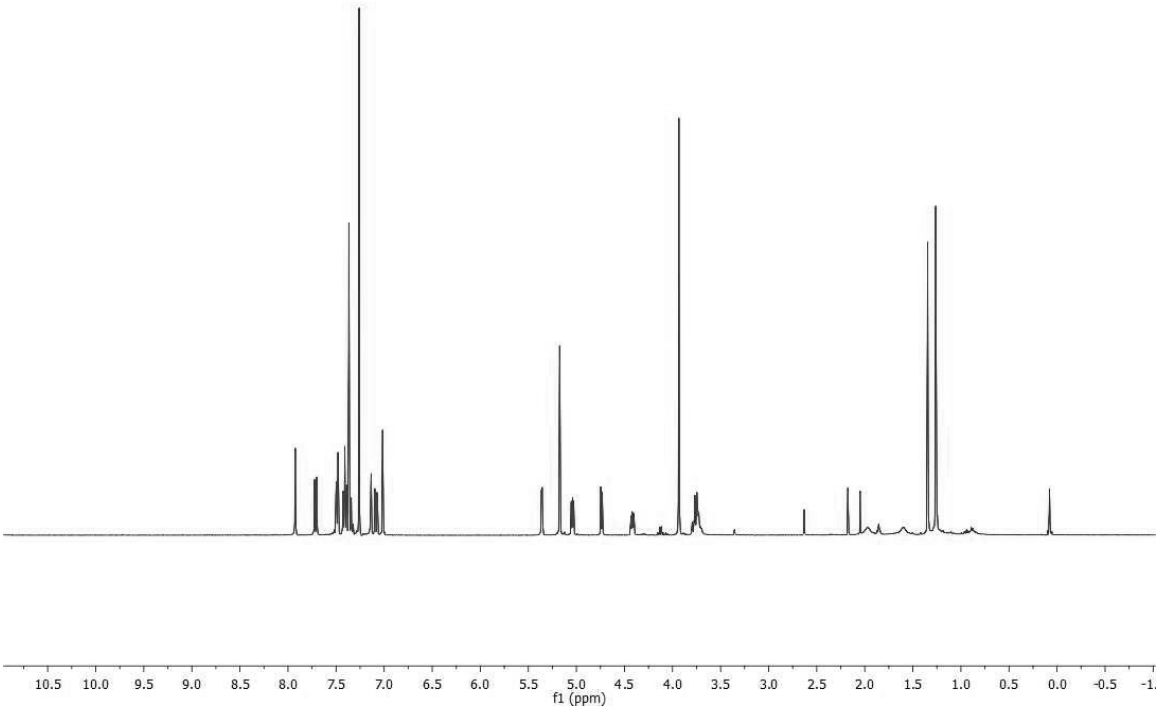


Figure S5: 400 MHz ^{1}H NMR spectra ($CDCl_3$) of **4**

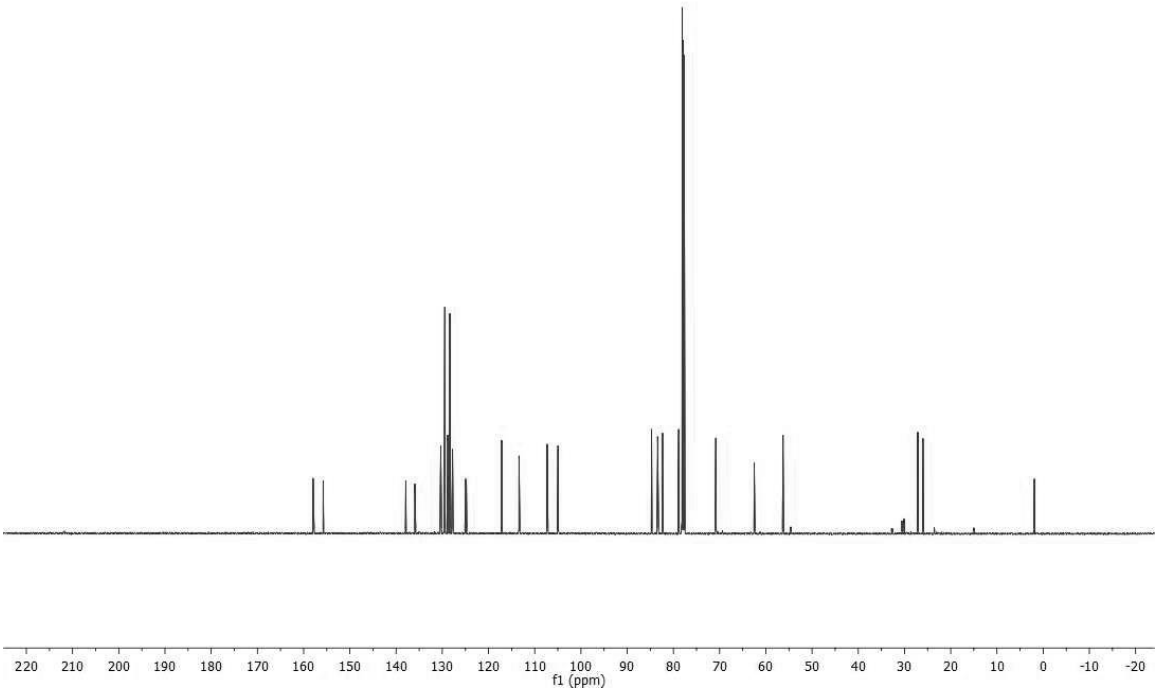


Figure S6: 151 MHz ^{13}C NMR spectra ($CDCl_3$) of **4**

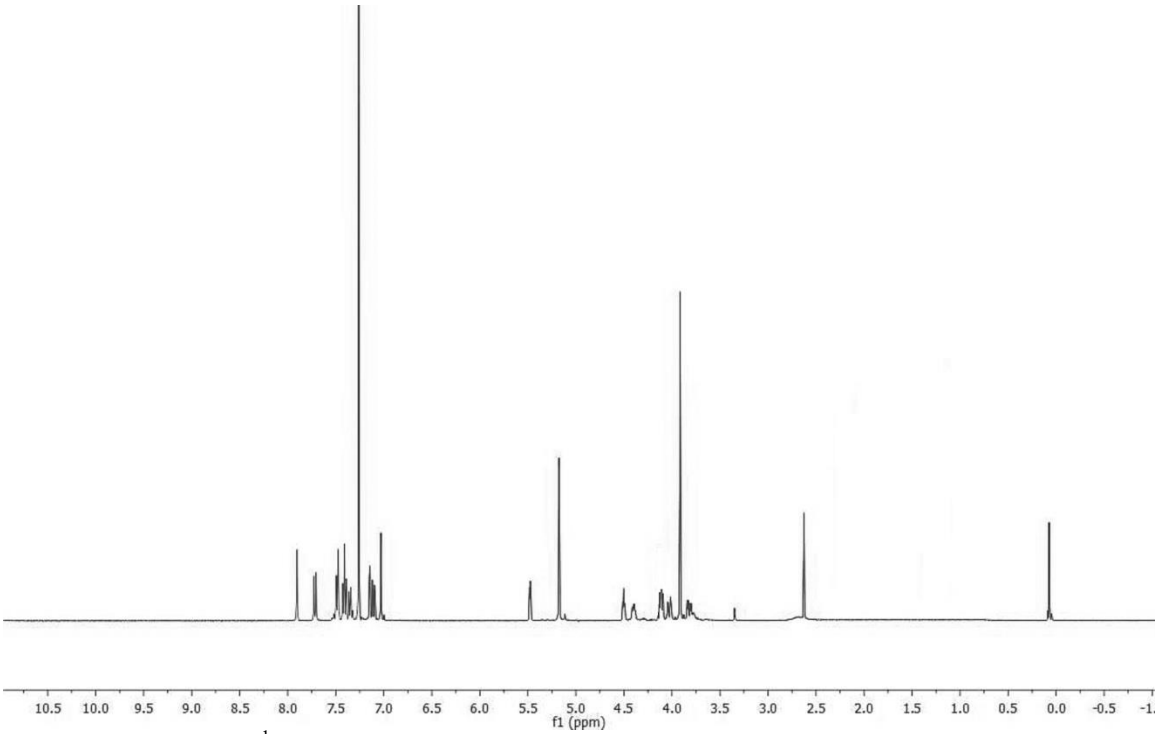


Figure S7: 400 MHz ^{1}H NMR spectra ($CDCl_3$) of **5**

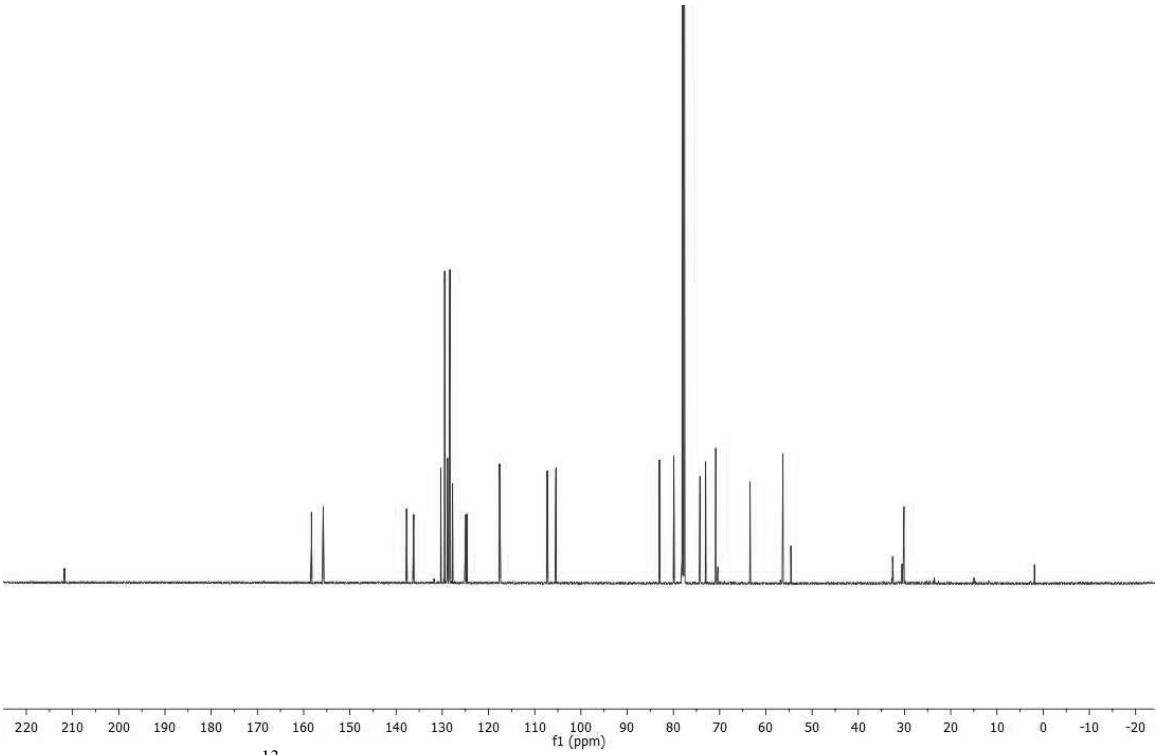


Figure S8: 151 MHz ^{13}C NMR spectra ($CDCl_3$) of **5**

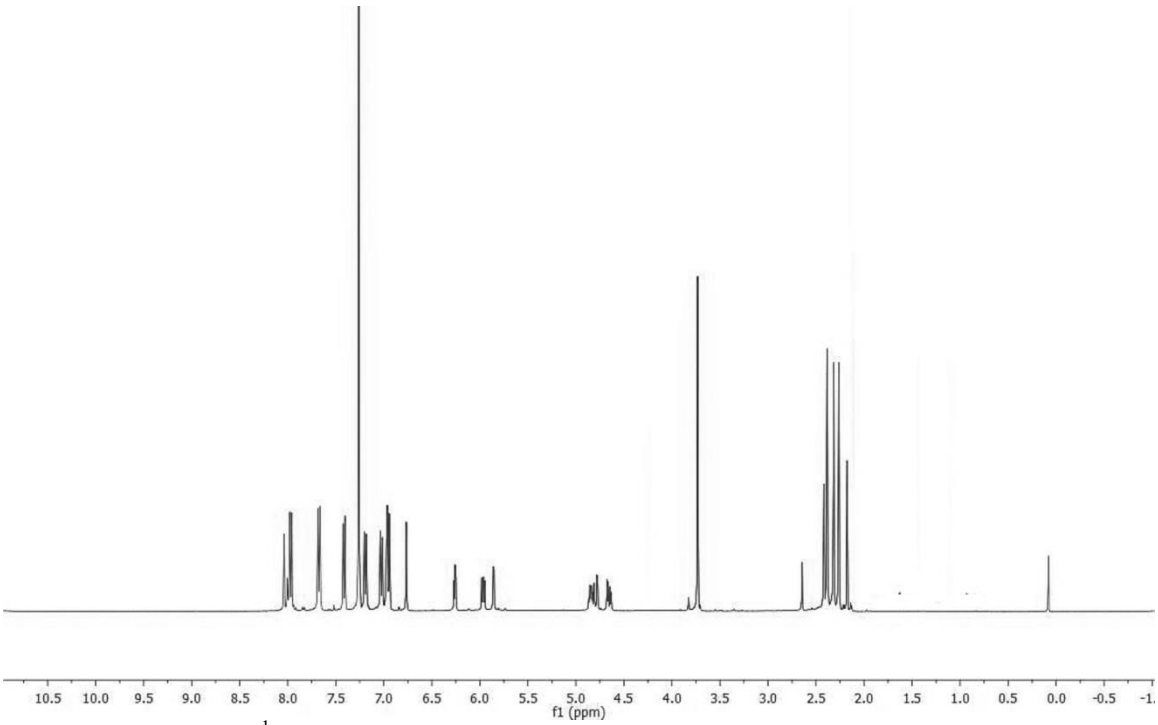


Figure S9: 400 MHz ^{1}H NMR spectra ($CDCl_3$) of **6**

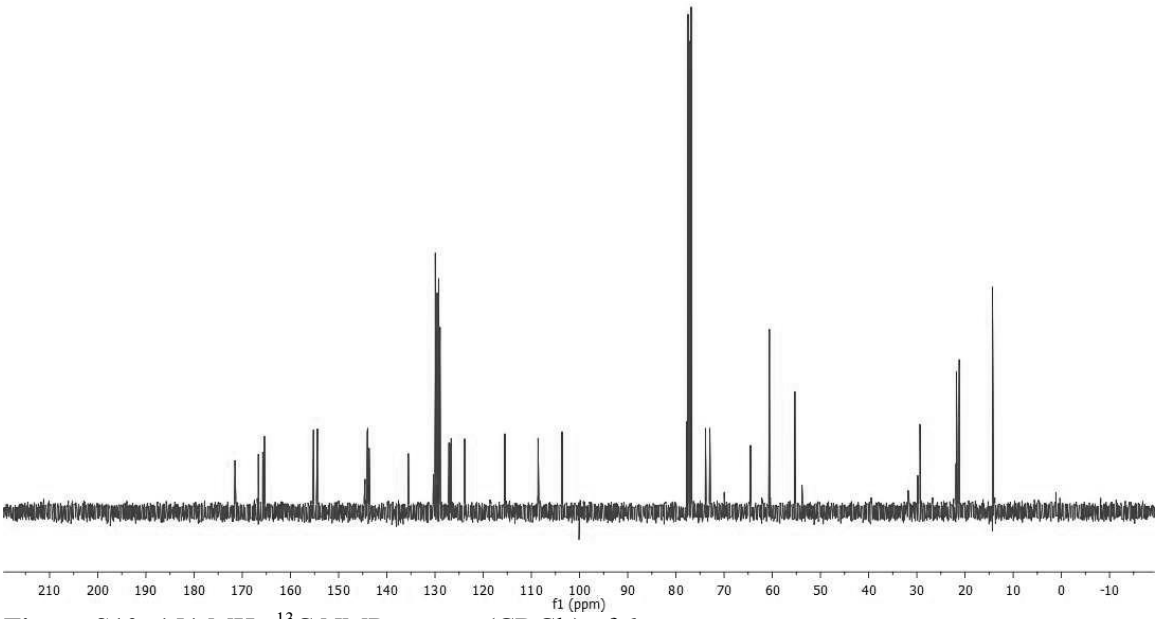


Figure S10: 151 MHz ^{13}C NMR spectra ($CDCl_3$) of **6**

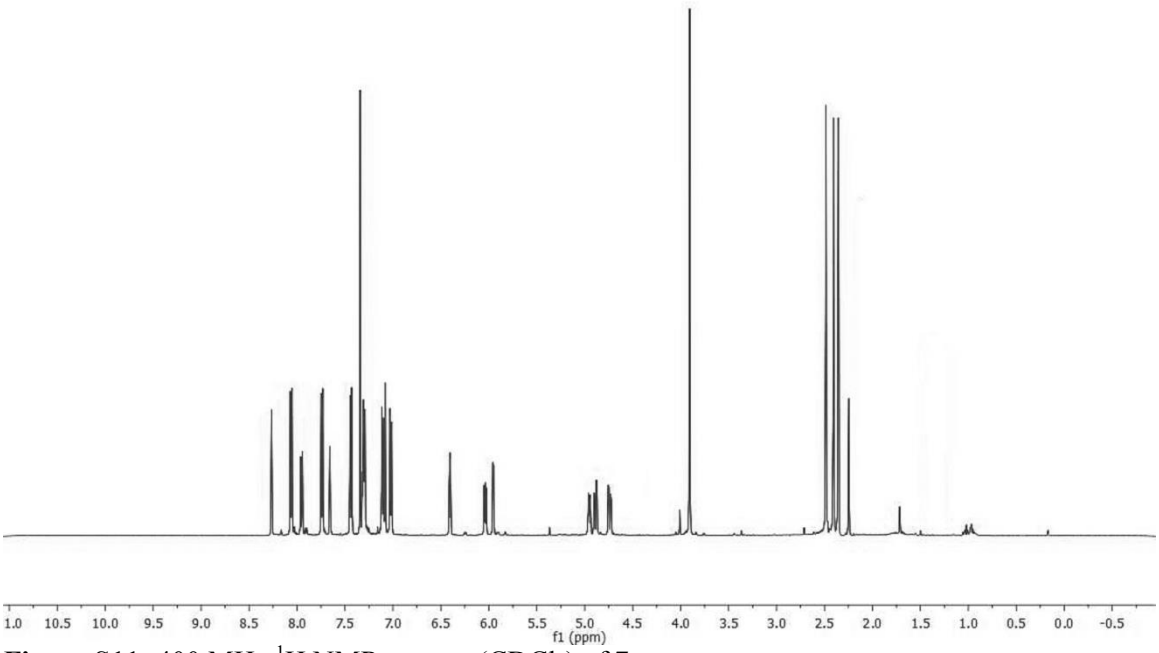


Figure S11: 400 MHz ^{1}H NMR spectra ($CDCl_3$) of **7**

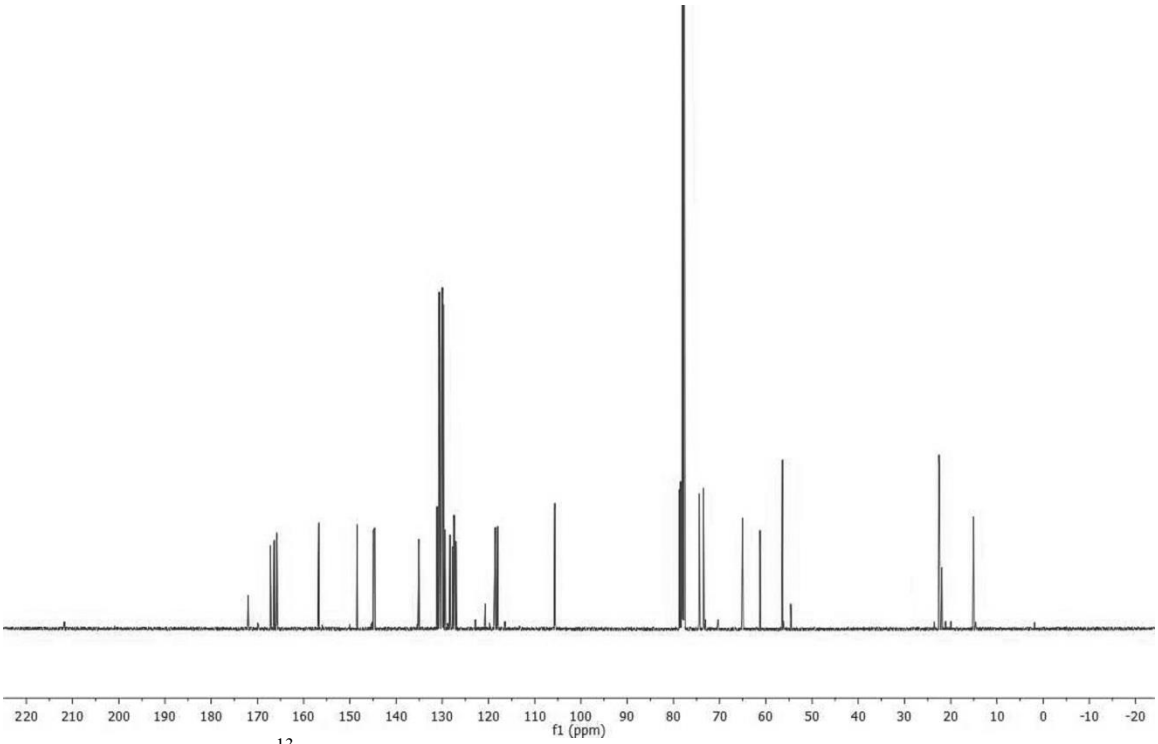


Figure S12: 151 MHz ^{13}C NMR spectra ($CDCl_3$) of **7**

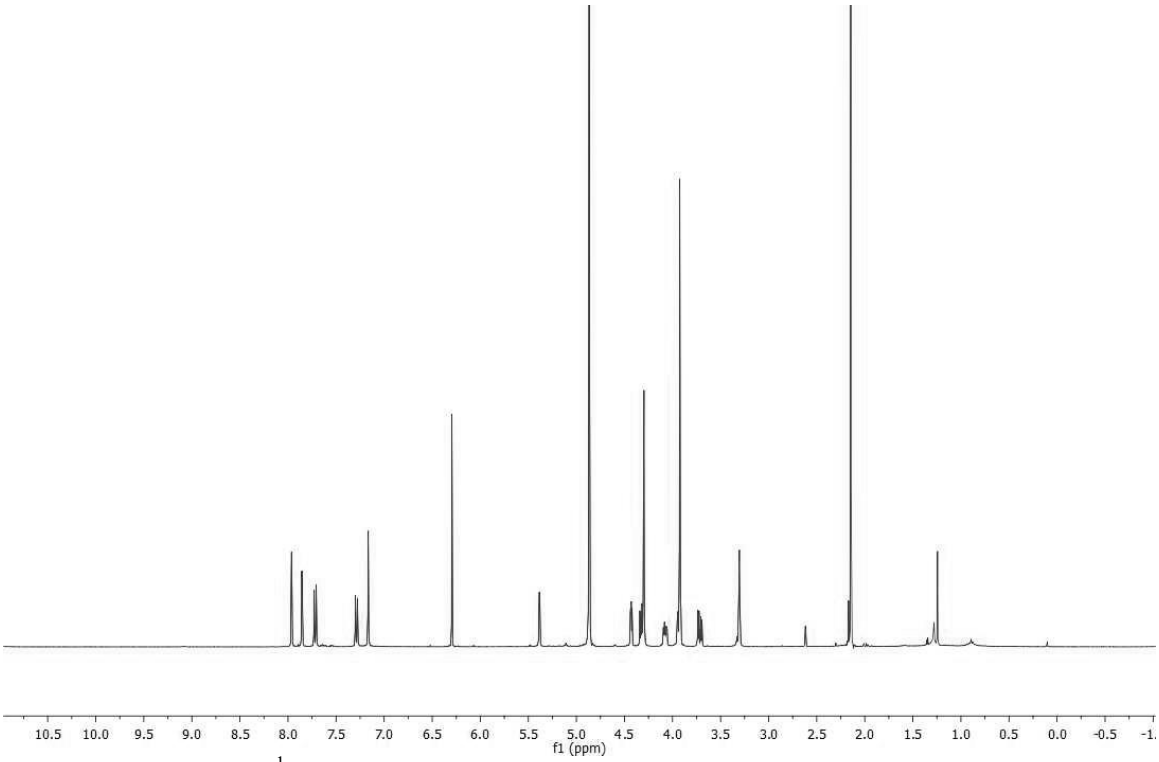


Figure S13: 400 MHz ^{1}H NMR spectra ($CDCl_3$) of **8**

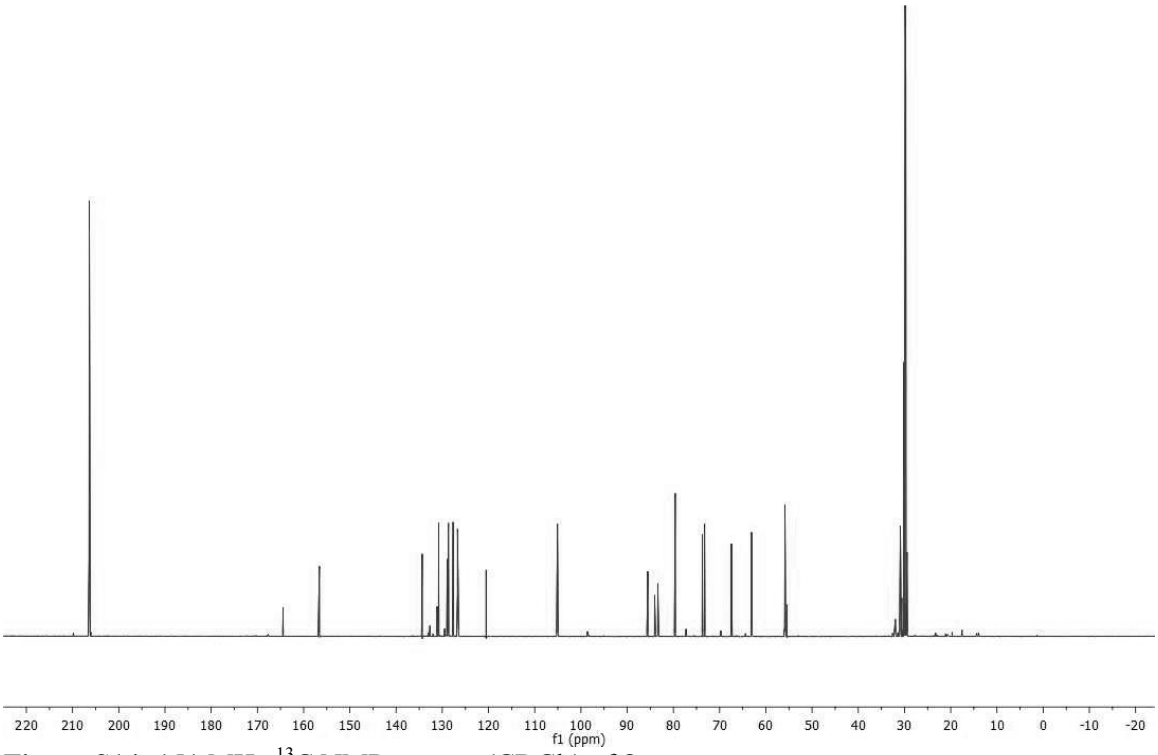


Figure S14: 151 MHz ^{13}C NMR spectra ($CDCl_3$) of **8**

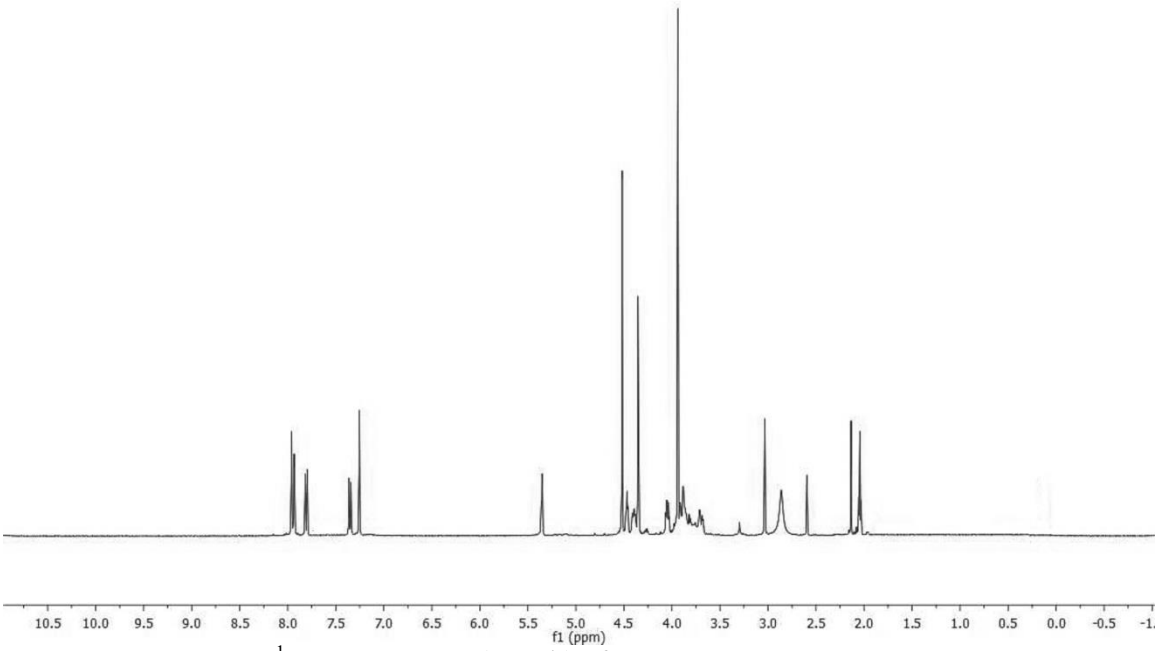


Figure S15: 400 MHz ^{1}H NMR spectra ($CDCl_3$) of **9**

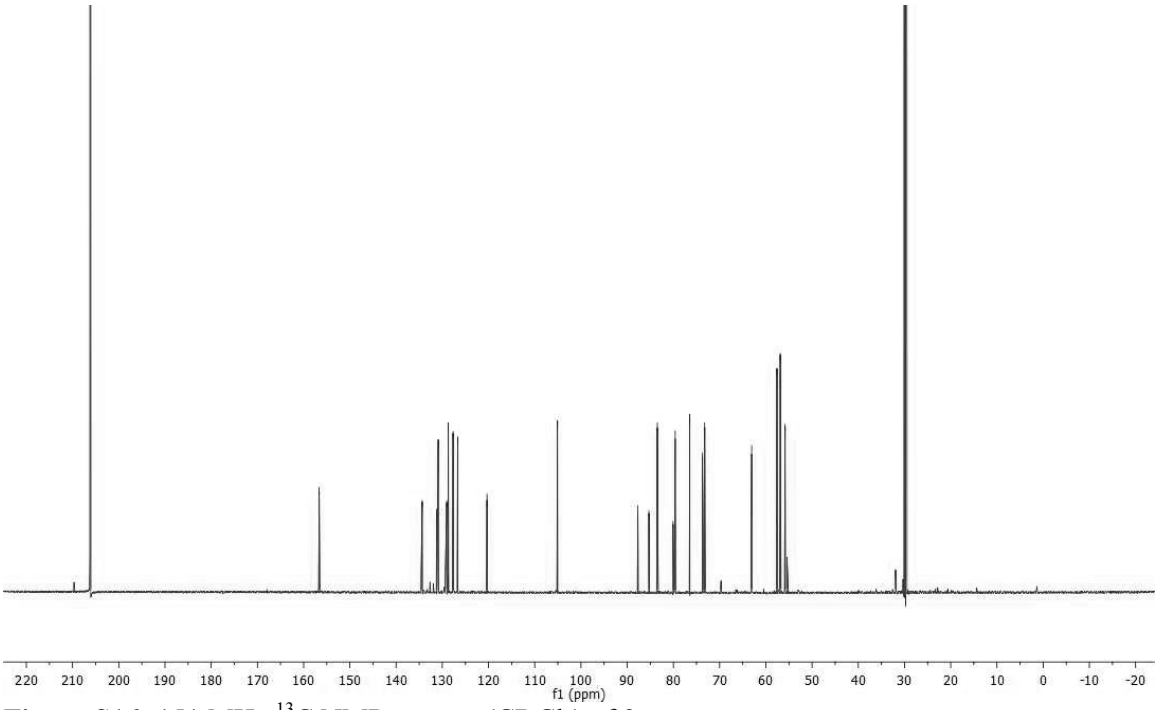


Figure S16: 151 MHz ^{13}C NMR spectra ($CDCl_3$) of **9**

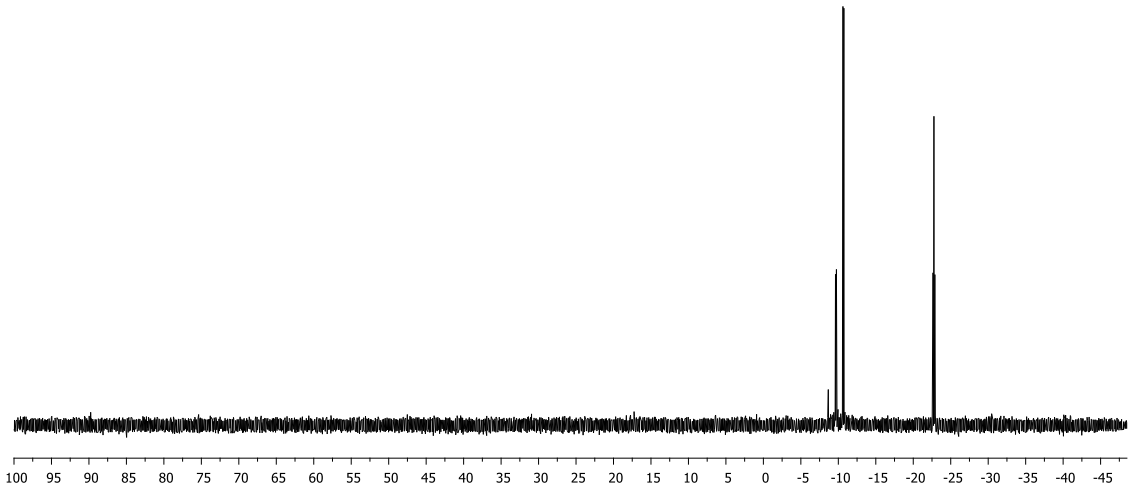


Figure S17: 162 MHz ^{31}P NMR spectra (D_2O) of **NaM^{A}**TP

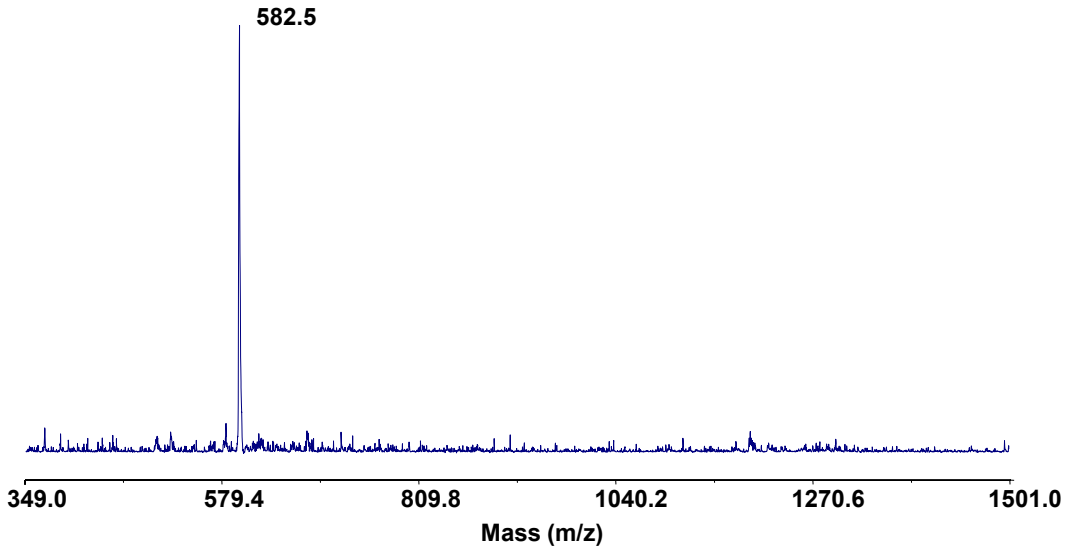


Figure S18: MALDI-TOF MS analysis of **NaM**$^{\mathbf{A}}$TP

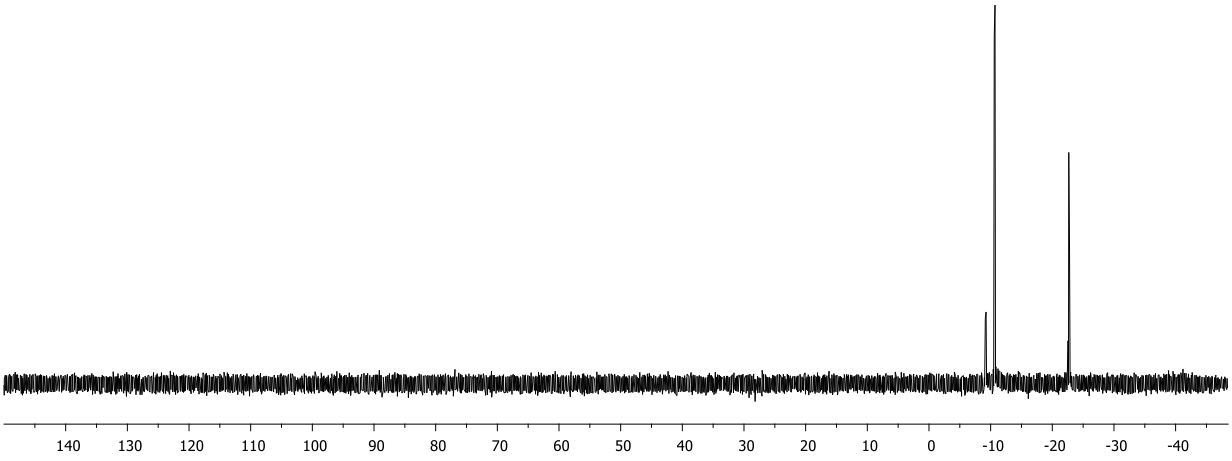


Figure S19: 162 MHz ^{31}P NMR spectra (D_2O) of **NaM**$^{\mathbf{CO}}$TP

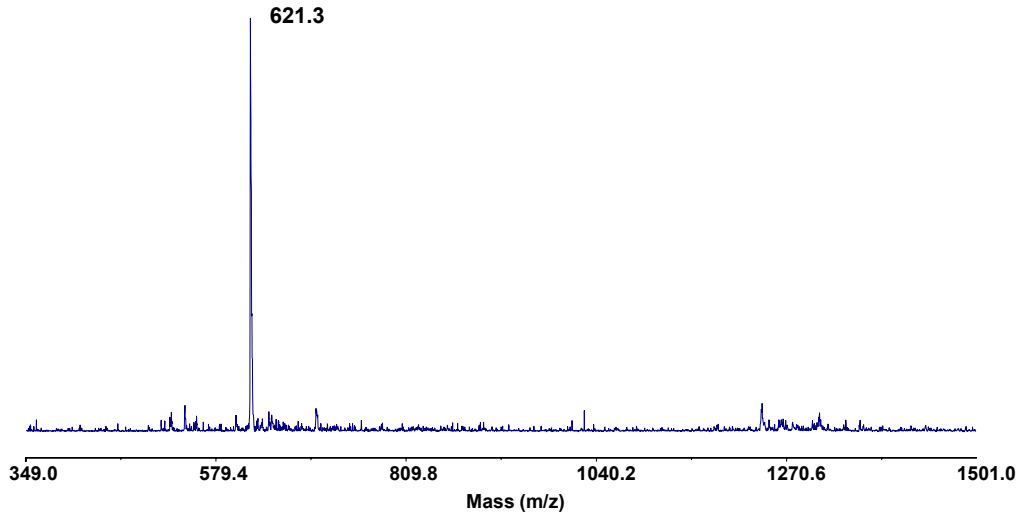


Figure S20: MALDI-TOF MS analysis of **NaM**^{CO}TP

DNA synthesis and PCR assembly

Materials. OneTaq and DeepVent DNA Polymerases were obtained from New England Biolabs. Sybr Gold (Life Technologies) was used for gel staining: PAGE gels were stained with 1× Sybr Gold for 30 min, agarose gels were cast with 1× Sybr Gold. Gels were visualized using a Molecular Imager Gel Doc XR+ equipped with 520DF30 filter (Bio-Rad) and quantified with Quantity One software (Bio-Rad). RNA was purified using RNA Clean and Concentrator-5 (Zymo Research, Irvine, CA) and eluted in 20 µL of RNAse-free water unless stated otherwise. The sequences of all DNA oligonucleotides used in this study are shown in Table S1. The concentration of dsDNA was measured by fluorescent dye binding (Quant-iT dsDNA HS Assay kit, Life Technologies). The concentration of RNA or ssDNA was determined by UV absorption at 260 nm using a NanoDrop 1000 (Thermo Scientific). d**NaM** phosphoramidite, d**NaM** and d**5SICS** nucleosides were obtained from Berry & Associates Inc (Dexter, MI). Free nucleosides of d**NaM** and d**5SICS** (Berry & Associates) were converted to the corresponding triphosphates under Ludwig conditions[3] as described.[4] After purification by anion exchange chromatography (DEAE Sephadex A-25) followed by reverse phase (C18) HPLC, both triphosphates were lyophilized and kept at -20 °C until use.

DNA synthesis. Natural oligonucleotides were purchased from IDT (San Diego, CA). Oligonucleotides containing an unnatural nucleotide were prepared using standard automated DNA synthesis methodology with ultra-mild DNA synthesis phosphoramidites on CPG ultramild supports (1 µmol, Glen Research; Sterling, VA) and an ABI Expedite 8905 synthesizer. After automated synthesis, the DMT-ON oligonucleotide was first purified by Glen-Pak cartridge (Glen Research) and then by 8 M urea 6% PAGE followed by "crush and soak" extraction with buffer (200 mM NaCl, 1 mM EDTA, 10 mM Tris pH 8) overnight, and finally desalted over Sephadex G-25 (NAP-10 Columns, GE Healthcare). The sequences of all DNA oligonucleotides are given in Table S1.

PCR assembly of the DNA templates. All PCR assembly steps (see Fig. S21) were carried out under the following general conditions: 1× OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.3 mM of dNTPs, 0.02 U/µL of OneTaq and 0.0025 U/µL of DeepVent DNA polymerases. Other specific conditions are given in Table S2. PCR reactions were divided into multiple aliquots 50 µL each with one followed in real time using 0.5× Sybr Green I (Life Technologies); following PCR, the aliquots were recombined and

purified by spin-column (DNA Clean and Concentrator-5; Zymo Research, Irvine, CA) followed by elution in 25 μL of water. One spin column was used to purify 4×50 μl of PCR reaction. For large scale transcription, the 260 bp template was purified by 4% agarose gel, recovered with Zymoclean Gel DNA Recovery Kit, and reamplified in 60×50 μL PCR reactions to generate ~10 μg of dsDNA. PCR amplicons were Sanger sequenced as described[5] to confirm the retention of the unnatural base pairs. Over 90% retention for each of the two positions was observed (Fig. S22).

Table S1. Oligonucleotide sequences used for 16S RNA central fragment assembly. T7 RNA promoter is shown in italics, priming regions are underlined

Template name	Oligonucleotide template (5' to 3') X = d**NaM**	PCR primers
T85	65mer : C CCG GGT ATC TAA TCC GGT TTG CTC CCC ACG CT X TCG CGC CTC AGC GTC ACG GGT GGA CCA GGT G	P3-short: C CCG GGT ATC TAA TCC G overlap-T7: *TAA TAC GAC TCA CTA TA* GGG C ACC TGG TCC ACC CGT GAC
T86	70mer: G TAG CGG TGA AAT GCG CAG ATA CCG GGA GGA ACG CC X ATG GCG AAG GCA GCC ACC TGG TCC ACC CGT GAC	P2-T7: *TAA TAC GAC TCA CTA TA* GGG TAG CGG TGA AAT GCG overlap-C: ACG GGT GGA CCA GGT G
T135	65mer 70mer	P2-T7 P3-short
Template260	P1 long : *TAA TAC GAC TCA CTA TA* GGG CGT GTA GGC GGC CTG GGG CGT CCC ATG TGA AAG ACC ACG GCT CAA CCG TGG GGG AGC GTG GGA TAC P2-reverse long CGC ATT TCA CCG CTA CTC CGG GAA TTC CAC CAC CCT CTC CCA CCG TCT AGC CTG AGC GTA TCC CAC GCT CCC CCA C 65mer 70mer	P1 : *TAA TAC GAC TCA CTA TA GGG C* P2-rev : CGC ATT TCA CCG CTA C P2 : G TAG CGG TGA AAT GCG P2-T7 P3-long : GTT TAG GGC GTG GAC TAC CCG GGT ATC TAA TCC G

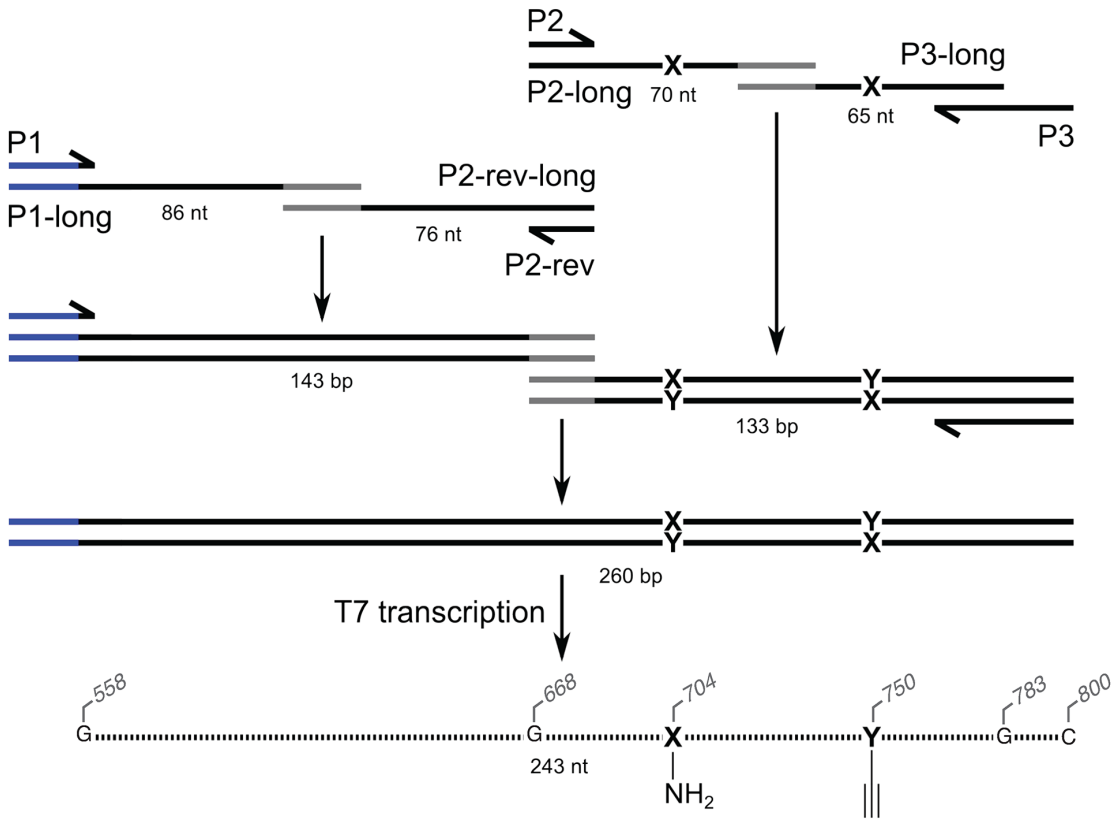


Figure S21. Assembly scheme for the 243-nt fragment of 16S ribosomal RNA containing functionalized unnatural **MMO2^{A}** (**X**) and **5SICS^{CO}** (**Y**) nucleotides. The numerical positions of **X** and **Y**, as well as the termini of the RNAs transcribed from the DNA templates T260 (nucleotides G558 to C800) and T135 (nucleotides G668 to G783) are indicated in gray. T7 promoter region is shown in blue. bp = base pair; nt = nucleotide.

Table S2. Specific PCR conditions for the DNA template assembly

Product	**143bp**	**133bp**	**260bp**
Templates	P1-long (4 ng) P2-rev-long (4 ng)	P2-long (4 ng) P3-long (4 ng)	143bp (4 ng) 133bp (4 ng)
d**NaM**TP, µM	-	100	100
d**5SICS**TP, µM	-	100	100
Primers (1 µM each)	P1, P2-rev	P2, P3	P1, P3
Volume, µl	200	200	200
Thermocycling conditions	Initial denaturation: 96 °C, 1 min 5 cycles: 96 °C, 10 s 68 °C, 30 s 10 cycles: 96 °C, 10 s 60 °C, 15 s 68 °C, 30 s	Initial denaturation: 96 °C, 1 min 10 cycles: 96 °C, 30 s 68 °C, 4 min 16 cycles: 96 °C, 30 s 60 °C, 30 s 68 °C, 4 min	Initial denaturation: 96 °C, 1 min 16 cycles: 96 °C, 30 s 60 °C, 30 s 68 °C, 4 min

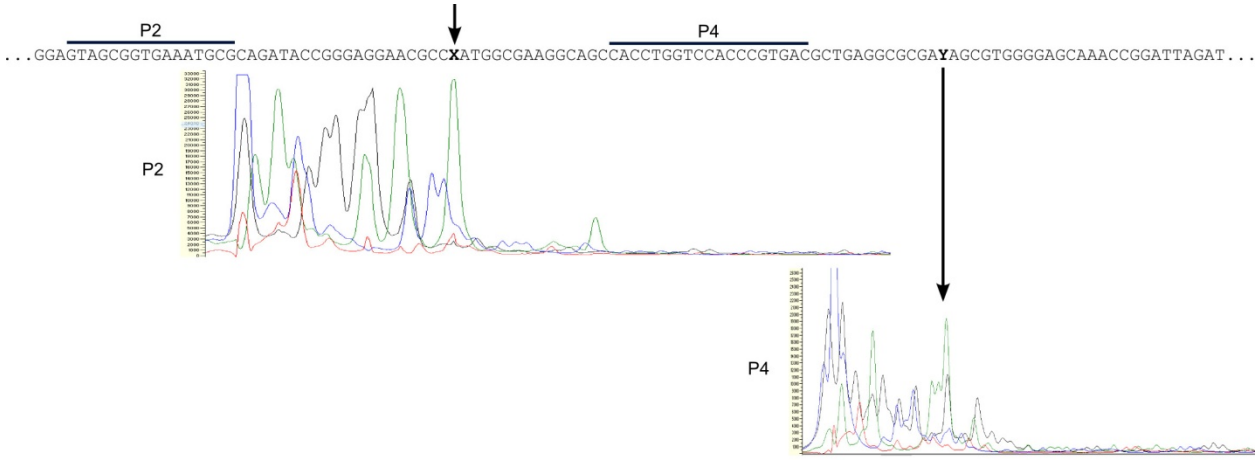
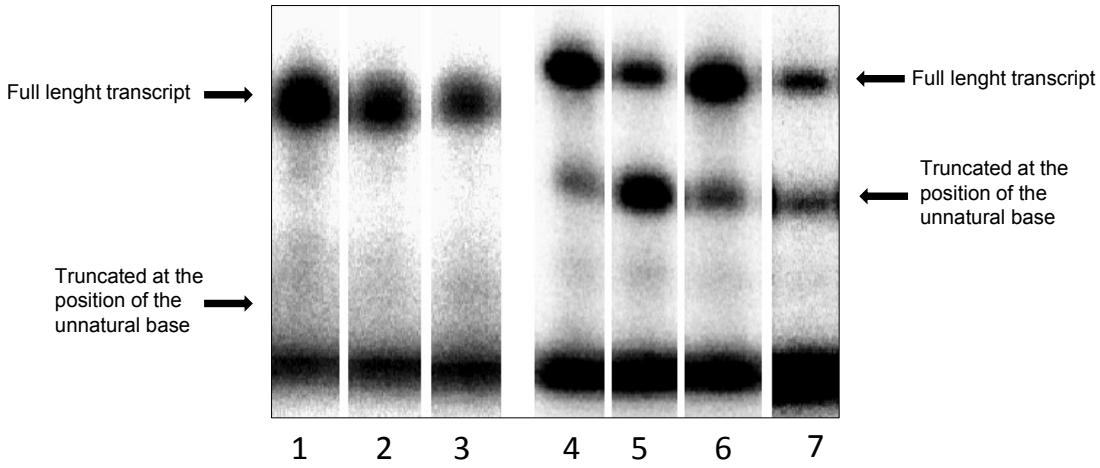


Figure S22. Raw sequencing traces with primers P2 and P4 (priming sequences are labeled with black boxes). **X** = d**NaM**; **Y** = d**5SICS**. The positions of the unnatural base pairs are shown with black arrows.

T7 transcription and RNA labeling

T7 transcription. Transcription of T85 and T86 dsDNA templates was carried out at 37 °C in 20 μL reactions containing 100 nM DNA (2 pmol), 1× Takara buffer, 16 mM $MgCl_2$, 5 mM DTT, 0.01% Triton X-100, 2 mM NTPs, [γ–P^{32}]GTP, 2.5 unit/μL T7 RNAP, and 1 mM of an unnatural ribotriphosphate. All solutions were prepared with DEPC-treated and nuclease-free sterilized water (Fisher Bioreagents). After incubation for 1 h at 37 °C, the reactions were quenched by addition of gel loading dye solution (20 μL of 10 M urea, 0.05% bromophenol blue). This mixture was heated at 75 °C for 3 min and then loaded on a 8 M urea 15% polyacrylamide gel using 1× TBE buffer. The gel was removed from the apparatus, and radioactivity was quantified by phosphorimaging (overnight exposure) using ImageQuant (Quantity One).



Lane description:

Lane	Template	Unnatural TP	Ratio full length/truncated product
1	T86	**5SICS**	87%
2	T86	**5SICS^{A}**	81%
3	T86	**5SICS^{CO}**	70%
4	T85	**NaM**	82%
5	T85	**NaM^{A}**	35%
6	T85	**MMO2^{A}**	75%
7	T85	**NaM^{CO}**	55%

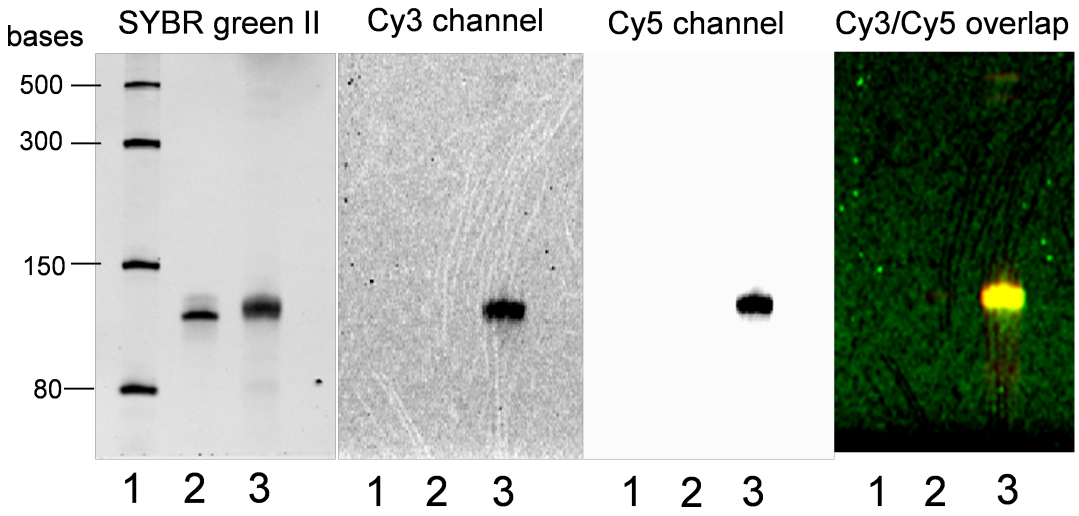
Figure S23. PAGE analysis of T85 and T86 transcription experiments

Transcription of T135, T260 and a fully natural analog of T260 (T260N) dsDNA templates was carried out at 37 °C in 100 μL reactions containing 100 nM DNA (10 pmol), 1× Takara buffer, 15 mM $MgCl_2$, 5 mM DTT, 0.01% Triton X-100, 2 mM NTPs, 1 mM **5SICS^{CO}**TP, 1mM **MMO2^{A}**TP [α–P^{32}]GTP and 2.5 unit/μL T7 RNAP. After incubation for 5 h at 37 °C, the reactions were quenched by addition of gel loading dye solution (100 μL of 10 M urea, 0.05% bromophenol blue). This mixture was heated at 75 °C for 3 min and then separated on a 8 M urea 6% polyacrylamide gel using 1× TBE buffer. Gels containing transcription products were imaged (overnight exposure with Kodak X-OmatAR5 film), and the image was used to identify the radioactive spots corresponding to the full-length [^{32}P]-RNA transcripts. These portions of the gel were then excised, transferred to microcentrifuge tubes and crushed in DEPC-treated water. The resulting mixtures were cooled to -80°C for 10 min then warmed up to 40°C for 10 min. Eluted full-length products were precipitated with cold ethanol (3 vol) in presence of CH_3COONa (0.3 M) at -20°C for 12 h. After centrifugation at 4 °C, the ethanol was removed, the pellets

were washed with CH_3COONa in EtOH (0,3M) and the products were evaporated to dryness with a SpeedVac concentrator. The purified full length RNAs were resuspended in DEPC treated water and subjected to post-transcriptional labeling.

RNA dual labeling with fluorophores. The purified full length RNA containing **MMO2^{A}** and **5SICS^{CO}** (transcript of T135 and T260) and the transcript of T260N were subjected to sequential labeling with sulfo-cyanine5 azide (Product# A3330, Lumiprobe, Hallandale Beach, FL) and Sulfo-cyanine3 NHS ester (Product# 11320, Lumiprobe).

Each RNA (3.5 μg, 8 μL) was mixed with 12 μL of degassed DMSO, heated at 85 °C for 5 min and cooled immediately on ice. The resulting solution was incubated with 4 μL of degassed Tris-HCl buffer (10×, pH 7.5), 4 μL of Sulfo-cyanine5 azide solution (10 mM in DMSO), 4 μL of $(BimC_4A)_3$ ligand (20 mM in DMSO, Sigma-Aldrich), 4 μL of CuSO4 (2 mM) and 4 μL of sodium ascorbate (25 mM) for 1.5 h at room temperature and the reaction was quenched by EDTA (20 μL, 5 mM). RNA were purified on RNA Clean & concentrator columns (Zymo) and eluted with 10 μL of DEPC treated H_2O. The resulting RNA (8 μL) were mixed with 12 μL of degassed DMSO heated at 85 °C for 5 min and cooled immediately on ice. The resulting solution was incubated with 6 μL of PBS buffer (5×, pH 7.4) and 4 μL of sulfo-cyanine3 NHS ester solution (10 mM in DMSO) for 1.5 h at room temperature. After purification on RNA Clean & concentrator columns (Zymo) the RNA were eluted in 10 μL of DEPC treated H_2O.



Lane description:

Lane	Sample	Dual labeling
1	Low-weight ssRNA ladder	-
2	T135 transcript	-
3	T135 transcript	+

Figure S24. PAGE analysis of T135 transcript dual labeling.

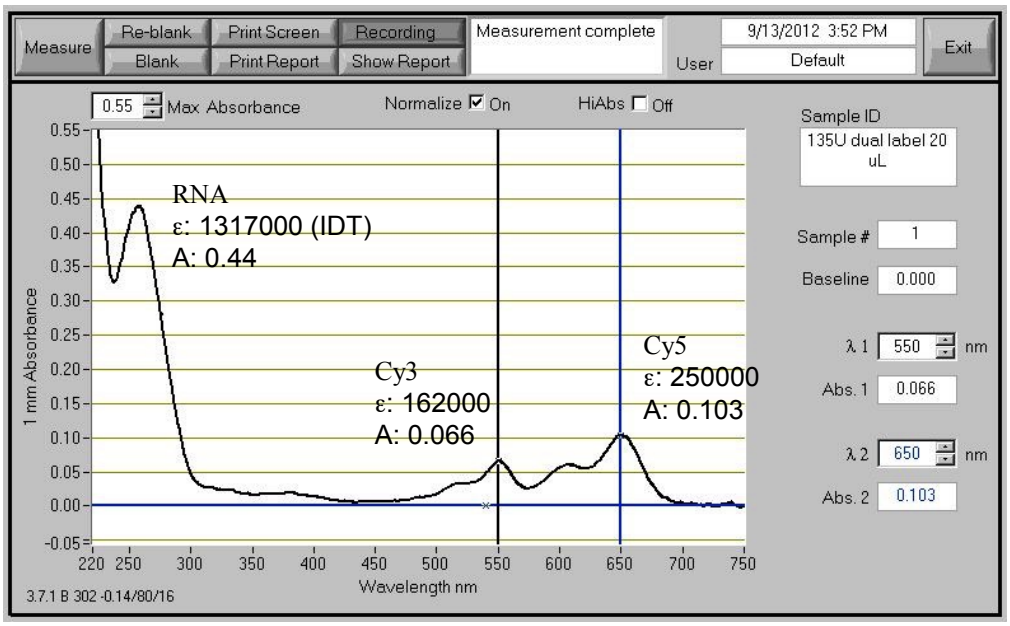
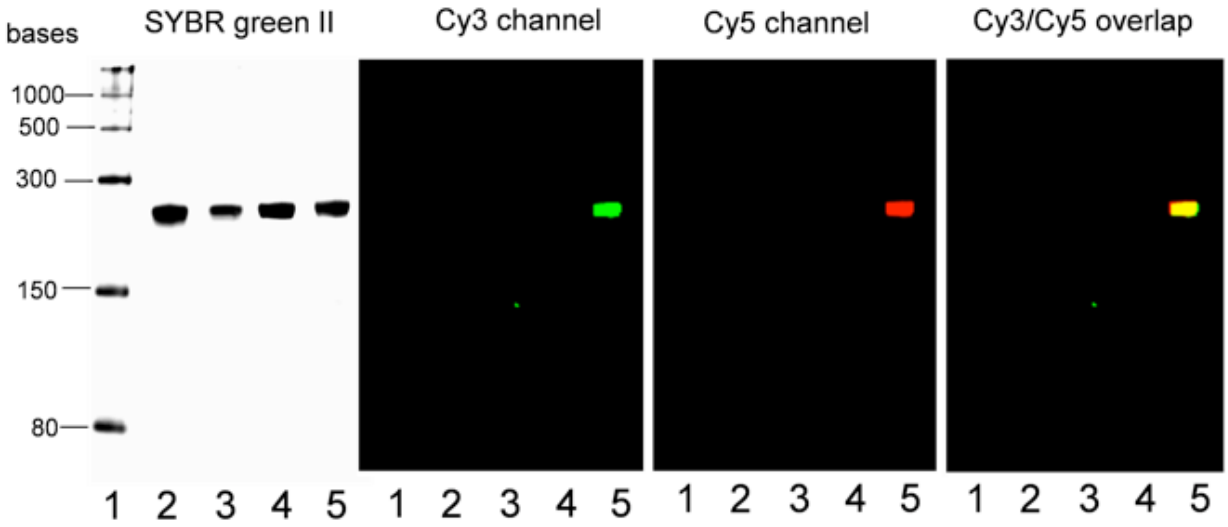


Figure S25. UV-VIS spectra of purified T135 transcript after dual labeling



Lane description

Lane	Sample	Dual labeling
1	Low-weight ssRNA ladder	-
2	T260N transcript	-
3	T260N transcript	+
4	T260 transcript	-
5	T260 transcript	+

Figure S26. PAGE analysis of T260N and T260 transcripts dual labeling.

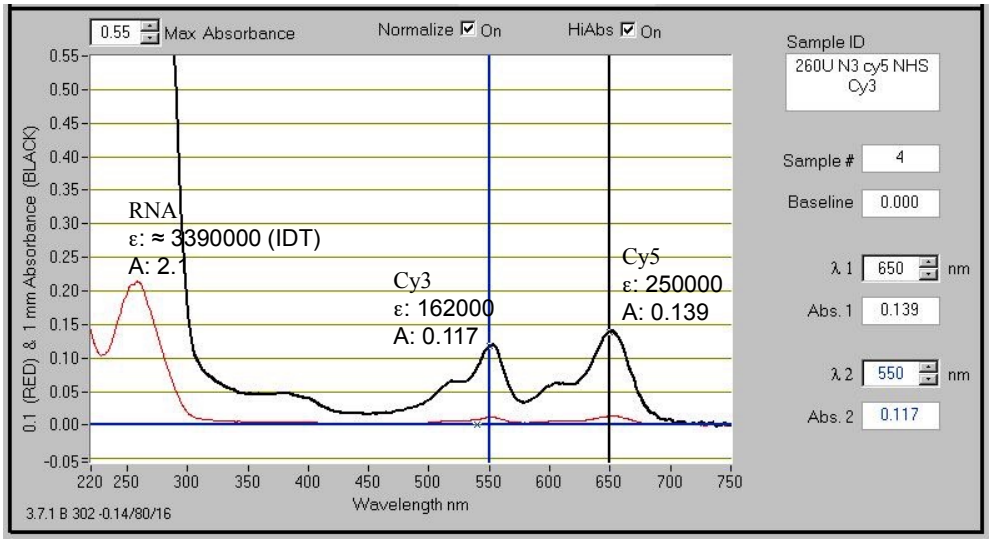


Figure S27. UV-VIS spectra of purified T260 transcript after dual labeling

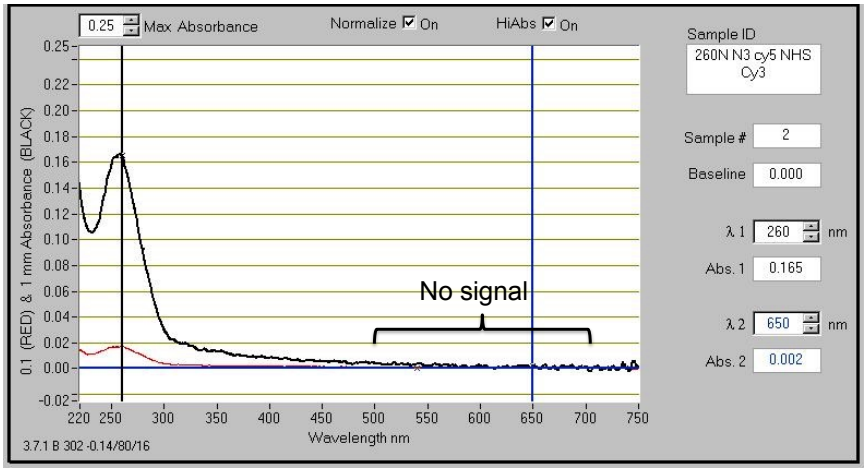
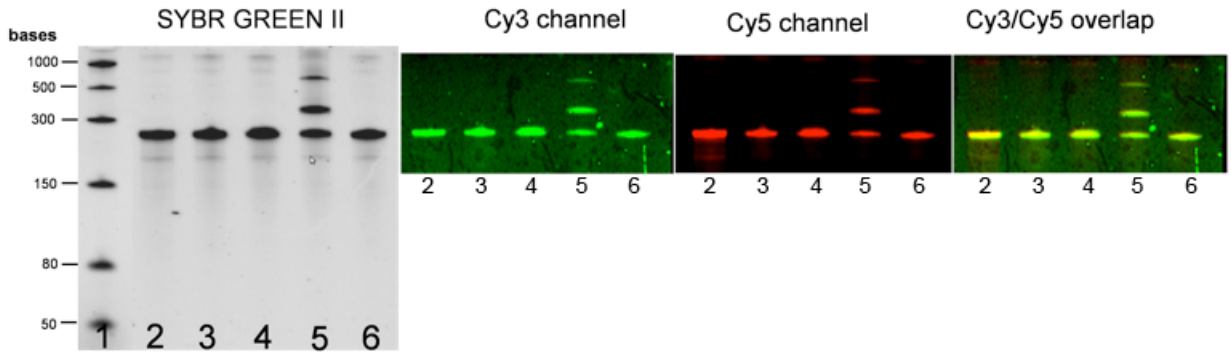


Figure S28. UV-VIS spectra of purified T260N transcript after dual labeling

3' RNA biotinylation. Dual labeled RNA was 3' biotinylated as described previously.[6] Briefly, RNA (80 µL, 50 ng/µL) was incubated with 10 mM $NaIO_4$ in 100 mM NaOAc buffer (pH 5.5) on ice for 1 h protected from light in a total volume of 100 µL. After clean up by RNA Clean and Concentrator-5 and elution in 20 µL of water, 2 µL were saved for quantification and PAGE analysis (Fig. S29), while the remaining amount was biotinylated with 5 mM Hydrazide-LC-Biotin (MW 371.50, Thermo Scientific) in 100 mM NaOAc buffer (pH 5.5) at 4°C overnight on ice in a total volume of 100 µl. After clean up by RNA Clean and Concentrator-5 and elution in 20 of water, 2 µl were saved for quantification and PAGE analysis (Fig. S29), while the remaining amount was used in single-molecule total internal reflection

fluorescence microscopy experiments. A mock oxidation and labeling was carried out with Cy3 and Cy5 dyes and their UV-vis spectra were monitored throughout the process to confirm their sufficient stability during the labeling.



Lane description:

Lane	Sample	Oxidation with $NaIO_4$	Reaction with biotin hydrazine	Streptavidin	% biotinylation based on: SYBR	Cy3	Cy5
1	low-weight ssRNA ladder						
2	Cy5/Cy3 dual labeled 16S 243 nt fragment	–	–	–			
3		+	–	–			
4		+	+	–			
5		+	+	+	53 %	55 %	54 %
6		+	–	+			

Figure S29. 3' biotinylation of Cy5/Cy3 dual labeled 16S fragment.

Expression and purification of ribosomal proteins

Wild type ribosomal S6, S15 and S18 proteins were expressed in E. coli strain BL21(DE3) carrying the pET22b(+) vector. DNA encoding the individual proteins was amplified from genomic DNA isolated from *Thermus thermophilus* strain HB8 (ATCC® 27634™) and inserted into the plasmid using the *Nde*I and *Hind*III sites.

Clarified cell lysates were subjected to a short heat exposure (15 min at 75 °C) followed by centrifugation, purifying each protein to near homogeneity (no other protein species detected by SDS PAGE). A second purification step on a HiTrap SP cation exchange column (GE Healthcare) was used to remove bound nucleic acids.

Single-molecule data acquisition and analysis

Single-molecule fluorescence data were acquired using a custom built prism-based TIRF microscope under oxygen scavenging conditions, as described.[7] Quartz slides were passivated with polyethylene glycol and coated with streptavidin before immobilizing the RNA sample, as described.[8] A Cy3 (donor) and Cy5 (acceptor) labeled RNA sample with biotin at the 3' end was heated to 90 °C and slowly cooled

to room temperature before immobilizing on the slide. A 100-150 pM RNA aliquot in imaging buffer (20 mM HEPES 7.5, 330 mM KCl, 1 mM DTT, and 10 mM $MgCl_2$ in 2 mM trolox) was introduced into the sample chamber and allowed to bind to the streptavidin-coated surface. Unbound RNA molecules were washed away and the imaging buffer was enriched with glucose oxidase and catalase as an oxygen scavenging system before data acquisition. RNA-protein (RNP) complexes were formed in imaging buffer at 65 °C for 10 minutes,[9] cooled to room temperature and mixed with the oxygen scavenging system before injection into the sample chamber. Immobilized RNA or RNP complexes were excited using a 532 nm laser and the emission from both donor and acceptor probes were spectrally separated and simultaneously recorded on different segments of an intensified CCD camera (Andor Technology, Belfast U.K.) with 100 ms integration time. A single-molecule data acquisition package (downloaded from https://physics.illinois.edu/cplc/software/) was used to record camera data and generate fluorescence intensity traces for both donor and acceptor. Individual intensity trajectories for the donor and acceptor channels were each corrected for background and were used to calculate FRET efficiency trajectories, using the formula $E = I_A / (\gamma I_D + I_A)$, where E is the FRET efficiency and I_D and I_A are the donor and acceptor intensities, respectively, and γ is a correction factor that accounts for differences in quantum yield and detection efficiency between donor and acceptor. Since γ was set to unity, the reported FRET efficiencies are apparent rather than absolute values. All data was processed using custom code written in MATLAB, as described previously.[10] The intensity traces showing anti-correlated fluctuations of the donor and acceptor emission as well as a single-step photobleaching transition were selected for further analysis after 7-point smoothing. The smoothed FRET traces were binned to generate a FRET histogram for each analyzed FRET trace and a composite FRET histogram was then compiled from multiple traces, using IGOR Pro (WaveMetrics). Individual peaks in the FRET histograms were fitted with Gaussian functions, using IGOR Pro software.

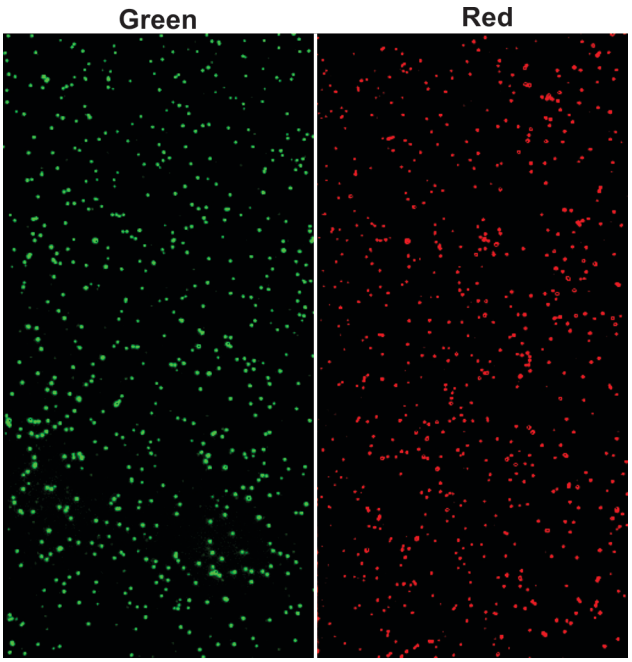


Figure S30. TIRF images of a field of immobilized RNA molecules recorded in green (Cy3) and red (Cy5) emission channels (left and right, respectively). The surface-immobilized RNA molecules were simultaneously excited with green (532 nm) and red (643 nm) lasers. Each of the spots in the green or red images represents individual immobilized RNA molecules. Custom software was used to match a spot in the green image with the corresponding spot in the red image, if present.

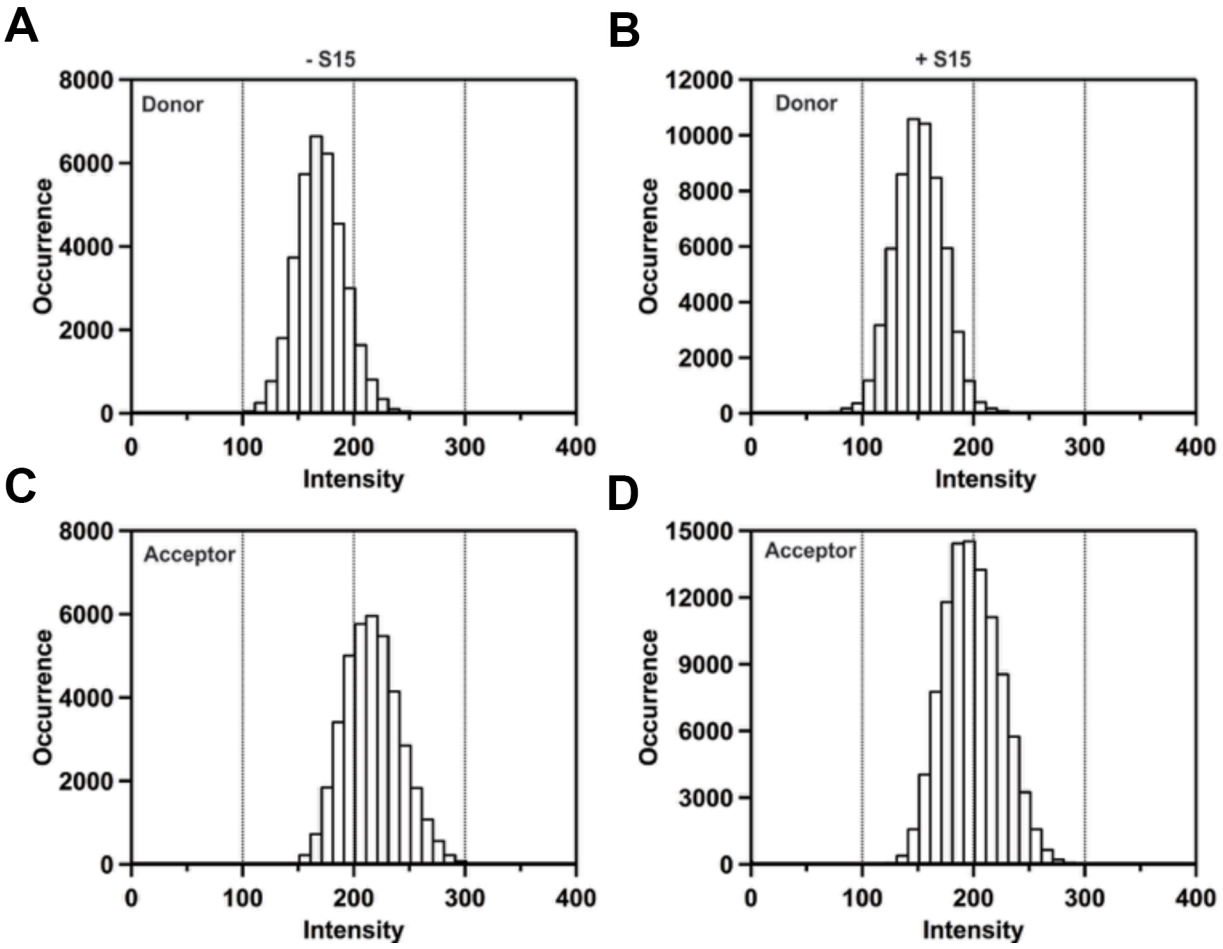


Figure S31. Fluorescence intensity histograms of Cy3 (donor) and Cy5 (acceptor) observed after direct excitation of Cy3 or Cy5. (A) Intensity histogram of Cy3 donor in the absence of S15, compiled from 55 individual doubly labeled RNA molecules. Cy3 was excited at 532 nm. (B) Intensity histogram of Cy3 in the presence of S15 protein (20 nM), showing no change in the mean donor intensity compared with panel (A). C) Intensity histogram of Cy5 acceptor in the absence of S15, compiled from 51 individual RNA molecules. Cy5 was directly excited at 647 nm. (D) Intensity histogram of Cy5, excited at 647 nm, in the presence of S15, showing little change in the mean acceptor intensity compared with panel (C). These results confirm that the S15 protein has no direct effect on the emission properties of the donor or acceptor.

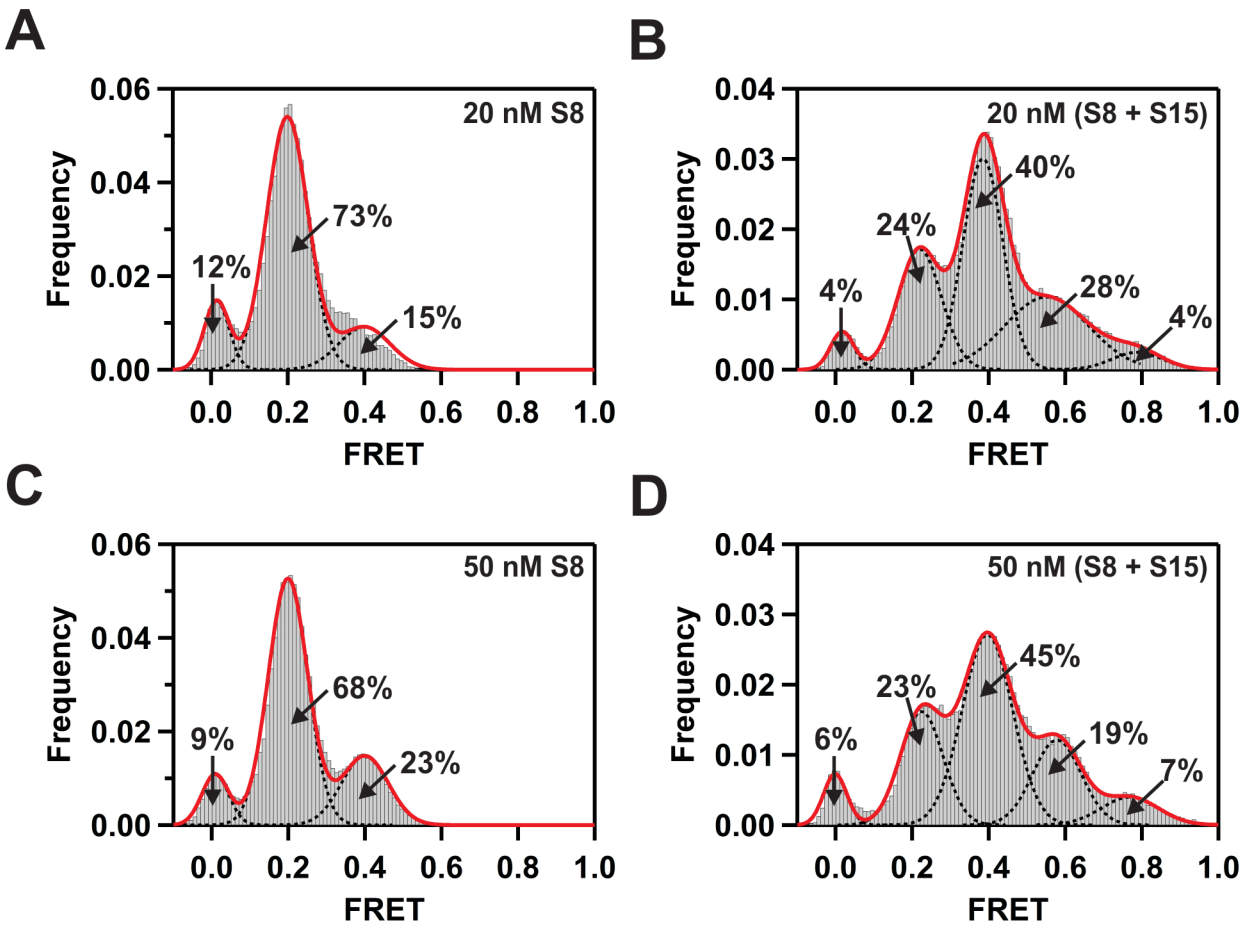


Figure S32. FRET efficiency histograms of doubly labeled RNA obtained under various experimental conditions. (A) In the presence of 20 nM of the ribosomal protein S8. (B) In the presence of both S8 and S15 (each 20 nM). (C) In the presence of S8 alone (50 nM). (D) In the presence of S8 and S15 (each 50 nM). In all cases, the dashed black lines are Gaussian fits to individual peaks and the red lines are the composite fit to the overall histogram.

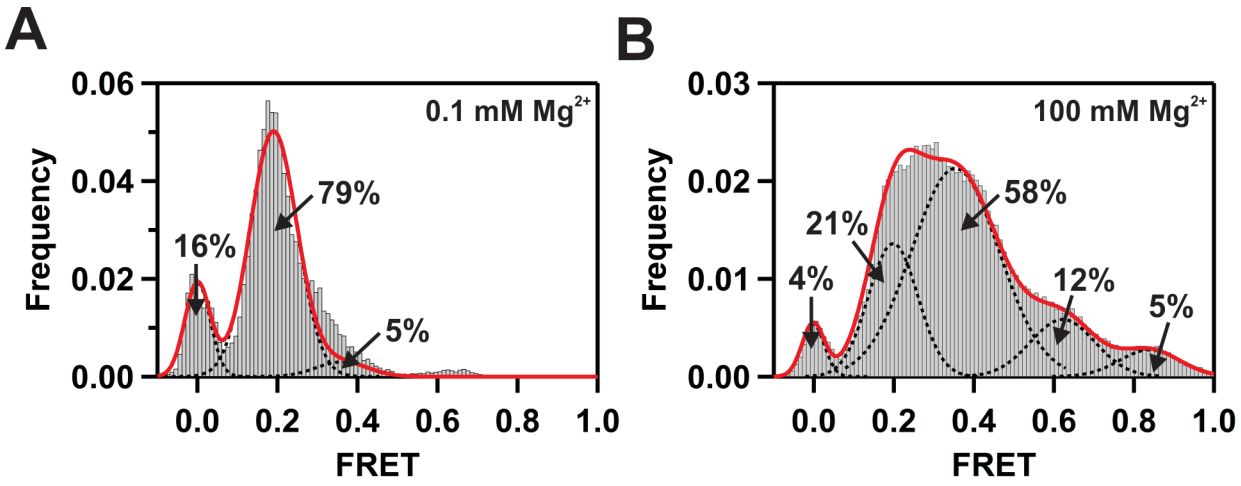


Figure S33. FRET efficiency histograms of doubly labeled RNA under various experimental conditions. (A) In the presence of 0.1 mM Mg^{2+} ions. (B) In the presence of 100 mM Mg^{2+} ions. In both (A) and (B), the dashed black lines are Gaussian fits to individual peaks and the red lines are the composite fit to the overall histogram.

References

1. Seo, Y. J., Malyshev, D. A., Lavergne, T., Ordoukhanian, P., and Romesberg, F. E. (2011) Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs, *J. Am. Chem. Soc. 133*, 19878–19888.

2. Anderson, S., Neidlein, U., Diederich, F., and Gramlich, V. (1995) Eine neue Klasse chiraler, von 1,1′-Binaphthyl abgeleiteter Cyclophan-Rezeptoren: Komplexierung von Pyranosiden, *Angew. Chem. 107*, 1722–1725.

3. Ludwig, J., and Eckstein, F. (1989) Rapid and efficient synthesis of nucleoside 5'-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one, *J. Org. Chem. 54*, 631–635.

4. Lavergne, T., Malyshev, D. A., and Romesberg, F. E. (2012) Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs, *Chem. Eur. J. 18*, 1231–1239.

5. Malyshev, D. A., Dhami, K., Quach, H. T., Lavergne, T., Ordoukhanian, P., Torkamani, A., and Romesberg, F. E. (2012) Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet, *Proc. Natl. Acad. Sci. USA 109*, 12005–12010.

6. Zearfoss, N. R., and Ryder, S. P. (2012) End-labeling oligonucleotides with chemical tags after synthesis, *Methods Mol. Biol. 941*, 181–193.

7. Berezhna, S. Y., Gill, J. P., Lamichhane, R., and Millar, D. P. (2012) Single-molecule Forster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I, *J. Am. Chem. Soc. 134*, 11261–11268.

8. Lamichhane, R., Solem, A., Black, W., and Rueda, D. (2010) Single-molecule FRET of protein-nucleic acid and protein-protein complexes: surface passivation and immobilization, *Methods 52*, 192–200.

9. Recht, M. I., and Williamson, J. R. (2004) RNA tertiary structure and cooperative assembly of a large ribonucleoprotein complex, *J. Mol. Biol. 344*, 395–407.

10. Zhuang, X., Kim, H., Pereira, M. J., Babcock, H. P., Walter, N. G., and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules, *Science 296*, 1473-1476.