Supplementary Material for Organic & Biomolecular Chemistry

Highly fluorescent nucleoside analog based on thieno[3,4-*d*]pyrimidine senses mismatched pairing

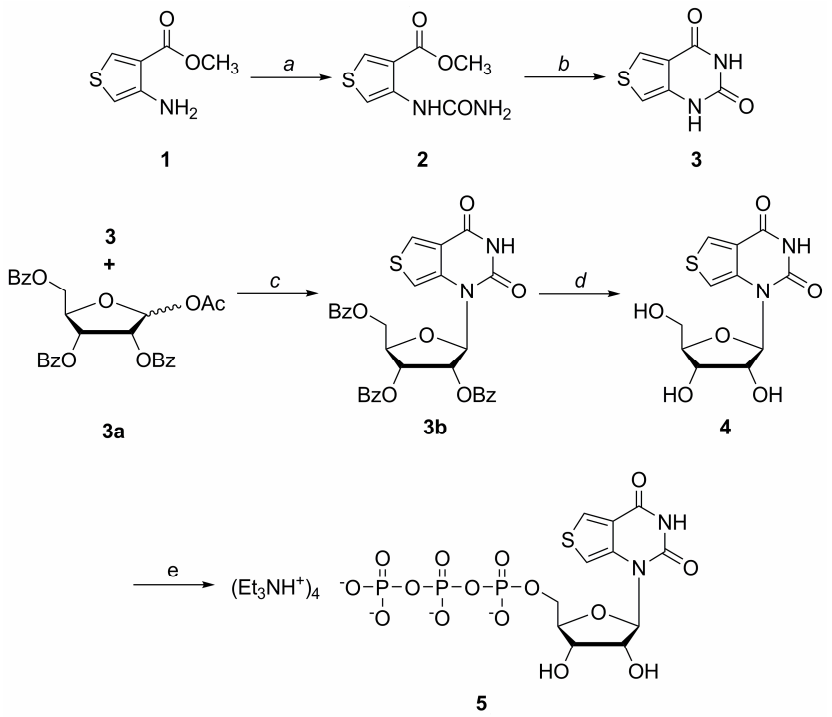
Seergazhi G. Srivatsan, Haim Weizman and Yitzhak Tor*

Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093-0358, USA. Fax: 858 534 0202; Tel: 858 534 6401; E-mail: ytor@ucsd.edu

Supporting Information

Contents

Syntheses



Scheme 1. Synthesis of thieno[3,4-*d*]pyrimidine-based emissive nucleoside and its triphosphate. *Reagents*: (*a*) KOCN, aq. acetic acid, RT, 81%; (*b*) NaOMe, MeOH, RT; (*c*) N-O-Bis(trimethylsilyl) acetamide, TMSOTf, CH_3CN, RT, 57% (Overall for steps *b* and *c*); (*d*) NH_4OH, dioxane, 60°C, 84%; (*e*) (i) $POCl_3$, $(MeO)_3PO$, 0−4°C; (ii) tributylammonium pyrophosphate, Bu_3N, 0−4°C, 56%.

Compound 2: To a solution of methyl 3-aminothiophene-4-carboxylate hydrochloride **1** (4.0 g, 20.7 mmol, 1 eq.) in acetic acid:water (1:1, 70 ml) was added KOCN (3.35 g, 41.3 mmol, 2.0 eq.) dissolved in water (30 ml) dropwise. The reaction mixture was stirred at RT for ~20 h and the precipitated product was filtered, washed with water (300 ml) and dried under vacuum (3.36 g, 81%). TLC (CH_2Cl_2:MeOH = 95:5) R_f = 0.43; ^{1}H-NMR (400 MHz, d_6-DMSO): δ 8.94 (br, 1H), 8.32 (dd, J_1 = 3.6 Hz, J_2 = 1.6 Hz, 1H), 7.69 (dd, J_1 = 3.6 Hz, J_2 = 1.2 Hz, 1H), 6.57 (br, 2H), 3.87 (s, 3H); ESI-MS (*m/z*) Calculated for $C_7H_8N_2O_3S$ [M] 200.03, found $[M+H]^+$ = 200.90.

Compound 3b: To a suspension of urea **2** (3.35 g, 16.7 mmol, 1.0 eq.) in anhydrous methanol (40 ml) was added NaOMe (1.98 g, 36.7 mmol, 2.2 eq.) dissolved in methanol (30 ml). The reaction mixture was stirred for 15 h at RT. The precipitate was filtered, washed with methanol (200 ml) and dried under vacuum (3.13 g). The weight of the

precipitate corresponded to more than 100% yield and the product did not elute out when silica gel chromatographic purification was attempted. However, the product **3** was found to be both TLC and NMR pure. ^{1}H-NMR (400 MHz, d_6-DMSO/$D_2$0): δ 7.97 (d, J = 3.2 Hz, 1H), 6.64 (d, J = 3.2 Hz, 1H); ^{13}C-NMR (100 MHz, d_6-DMSO/D_2O): δ 162.3, 154.6, 140.4, 130.2, 122.9, 104.2; ESI-MS (*m/z*) Calculated for $C_6H_4N_2O_2S$ [M] 168.00, found $[M+H]^+$ = 169.11.

To a suspension of the crude product **3** (1.0 g, 5.95 mmol) in anhydrous CH_3CN was added *N,O*-bis(trimethylsilyl) acetamide (BSA) at RT. After 30 min, TMS-OTf was added and the suspension partially dissolved. The benzoyl-protected ribose **3a** (4.10 g, 8.13 mmol) was added in two portions (once in 12 h) by dissolving in CH_3CN (10 ml). In total the reaction was performed for 24 h. The reaction mixture was evaporated and the residue was dissolved in CH_2Cl_2 (150 ml), washed with saturated $NaHCO_3$ and brine. The organic extract was evaporated and purified by silica gel chromatography to afford the product **3b** as off white foam (overall yield, 1.87, 57%). TLC (CH_2Cl_2:MeOH = 95:5) R_f = 0.63; ^{1}H-NMR (400 MHz, $CDCl_3$): δ 9.07 (br, 1H), 8.23 (d, J = 3.2 Hz, 1H), 8.11−8.09 (m, 2H), 7.96−7.90 (m, 4H), 7.59−7.26 (m, 9 H), 7.02 (d, J = 3.2 Hz, 1H), 6.52 (d, J = 4.8 Hz, 1H), 6.19−6.07 (m, 2H), 4.94−4.90 (m, 1H), 4.77−4.68 (m, 2H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ 166.4, 165.7, 165.6, 157.7, 150.3, 136.6, 134.0, 133.9, 133.7, 131.5, 130.1, 130.1, 130.0, 129.7, 129.0, 128.9, 128.8, 128.8, 128.7, 123.2, 105.4, 89.2, 80.1, 71.7, 70.5, 63.8; ESI-MS (*m/z*) Calculated for $C_{32}H_{24}N_2O_9S$ [M] 612.12, found $[M+H]^+$ = 162.79, $[M+Na]^+$ = 635.05.

Compound 4: A solution of benzoyl-protected nucleoside **3b** (1.21 g, 1.98 mmol) in a pressure tube was dissolved in dioxane (15 ml) and aq. NH3 (30%, 80 ml). The solution was heated at 60 °C for 12 h. The pressure tube was then placed in crushed ice for 5 min and the contents were evaporated to dryness. The residue was purified by silica gel chromatography to afford the nucleoside as white solid (0.50 g, 84%). TLC (CH_2Cl_2:MeOH = 9:1) R_f = 0.24; ^{1}H-NMR (400 MHz, d_6-DMSO): δ 11.37 (br, 1H), 8.48 (dd, J_1 = 3.2 Hz, J_2 = 1.2 Hz, 1H), 7.65 (dd, J_1 = 3.2 Hz, J_2 = 1.2 Hz, 1H), 6.11 (d, J = 6.4 Hz, 1H), 5.25 (d, J = 6.0 Hz, 1H), 5.20 (t, J = 4.8 Hz, 1H), 5.09 (d, J = 5.2 Hz, 1H), 4.51−4.48 (m, 1H), 4.13−4.09 (m, 1H), 3.85 (d, J = 2.8 Hz, 1H), 3.68−3.67 (m, 2H); ^{13}C-

NMR (100 MHz, d_6-DMSO): δ 158.3, 151.3, 136.2, 130.8, 124.0, 107.4, 89.0, 85.7, 69.8, 68.8, 61.8; ESI-MS (*m/z*) Calculated for $C_{11}H_{12}N_2O_6S$ [M] 300.04, found $[M+H]^+$ = 300.97. λ_{max} = 260 nm (ε = 8210 $M^{-1}cm^{-1}$); λ_{max} = 304 nm (ε = 3650 $M^{-1}cm^{-1}$).

Compound 5: To an ice cold solution of nucleoside **4** (0.058 g, 0.19 mmol, 1 eq.) in trimethyl phosphate (1 mL) was added freshly distilled $POCl_3$ (53 μL, 0.58 mmol, 3.1 eq.). The solution was stirred for 25 h at 0−4 °C. A solution of *bis*-tributylammonium pyrophosphate (0.5 M in DMF, 2 ml, 1.0 mmol, 5.3 eq.) and tributyl amine (0.50 mL, 2.1 mmol, 11 eq.) was then rapidly added under ice-cold condition. The reaction was quenched after 30 min with 1 M triethylammonium bicarbonate buffer (TEAB, pH 7.5, 15 mL) and was extracted with ethyl acetate (20 mL). The aqueous layer was evaporated under vacuum. The residue was purified first on a DEAE sephadex-A25 anion exchange column (10 mM–1.0 M, TEAB buffer, pH 7.5) followed by reverse-phase HPLC (Vydac C18 column, 1.0 × 25 cm, 5 μm TP silica, 0–15% acetonitrile in 100 mM triethyl ammonium acetate buffer, pH 7.0, 30 min). Appropriate fractions were lyophilized to afford the desired triphosphate product (0.102 g, 56%). ^{1}H-NMR (400 MHz, D_2O): δ 8.20 (d, *J* = 3.2 Hz, 1H), 7.30 (d, *J* = 3.2 Hz, 1H), 6.08 (d, *J* = 6.8 Hz, 1H), 4.61 (t, *J* = 6.8 Hz, 1H), 4.35 (dd, J_1 = 6.8 Hz, J_2 = 4.8 Hz, 1H), 4.16–4.03 (m, 3H); ^{13}C-NMR (100 MHz, D_2O): δ 160.1, 151.9, 134.7, 131.5, 122.2, 108.1, 88.8, 83.2, 68.8, 68.6, 65.1; ^{31}P-NMR (162 MHz, D_2O): –9.27 (d, *J* = 19.4 Hz, P_γ), –10.63 (d, *J* = 19.4 Hz, P_α), –22.29 (t, J = 19.4 Hz P_β); ESI-MS (negative mode): Calculated for $C_{11}H_{15}N_2O_{15}P_3S$ [M] 539.94, found $[M–H]^-$ = 538.93.

Quantum yield determination: Quantum yield for nucleoside **4** relative to 2-aminopurine standard was determined using the following equation. (Du, H.; Fuh, R. A.; Li, J.; Corkan, A.; Lindsey, J. S. *Photochem. and Photobiol.,* **1998**, *68*, 141–142)

$$\Phi_{F(x)} = (A_s/A_x)\,(F_x/F_s)\,(n_x/n_s)^2\,\Phi_{F(s)}$$

Where s is the standard, x is the nucleoside, A is the absorbance at excitation wavelength, F is the area under the emission curve, *n* is the refractive index of the solvent and Φ_F is the quantum yield (Lavabre, D.; Fery-Forgues, S. *J. Chem. Educ.*, **1999**, *76*, 1260–1264).

Quenching Studies and Stern-Volmer plot: Quenching of nucleoside fluorescence by NMPs was performed by adding aliquots of concentrated nucleotide solution (20 mM) to the nucleoside **4**, where both solutions contained equal concentration of the nucleoside (5 μM). All solutions were prepared in HEPES buffer containing 0.5% DMSO (20 mM HEPES, 100 mM NaCl, 0.5 mM EDTA, pH 7.4). The nucleoside was excited at 304 nm with an excitation slit width of 5 nm and emission slit width of 8 nm, and changes in fluorescence was monitored at the emission maximum, 412 nm.

Stern-Volmer plot was obtained by plotting F_0/F vs. concentration of the quencher (NMPs) and was fit to equation 1.

$$F_0/F = 1 + K_{sv}[\text{NMPs}] \qquad (1)$$

Where F_0 and F are the fluorescence intensities in the absence and presence of quenchers (NMPs) respectively, and K_{sv} is the quenching constant.

Transcription reactions with α-^{32}P ATP: Single strand DNA templates were annealed to an 18-mer T7 RNA polymerase consensus promoter sequence in TE buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8) by heating a 1:1 mixture (5 μM) at 90°C for 3 min and cooling the solution slowly to room temperature. Transcription reactions were performed in 40 mM Tris-HCl buffer (pH 7.9) containing 250 nM annealed template, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 10 mM NaCl, 2 mM spermidine, 1 U/μL RNase inhibitor (RiboLock), 1 mM GTP, 1 mM CTP, 1 mM UTP, 1 mM modified UTP **5**, 20 μM ATP, 5 μCi α-^{32}P ATP and 2.5 U/μL T7 RNA polymerase (Fermentas) in a total volume of 20 μL. After 3 h at 37°C, reactions were quenched by adding 20 μL of loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, pH 8 and 0.05% bromophenol blue), heated to 75°C for 3 min, and samples (4 μL) were loaded onto an analytical 20% denaturing polyacrylamide gel. The products on the gel were analyzed using a phosphorimager.

Large-scale transcription reaction using template **6** was performed in 250 μL reaction volume under similar conditions to isolate RNA for enzymatic digestion. The reaction contained 2 mM NTPs, 2 mM modified UTP **5**, 20 mM $MgCl_2$ and 600 units T7

RNA polymerase. After incubation for 12 h at 37°C, the precipitated magnesium pyrophosphate was removed by centrifugation. The reaction volume was reduced to half by Speed Vac and 25 µL of loading buffer was added. The mixture was heated at 75°C for 3 min, and loaded onto a preparative 20% denaturing polyacrylamide gel. The gel was UV shadowed; appropriate band was excised, extracted with 0.5 M ammonium acetate and desalted on a Sep-Pak.

Enzymatic Digestion: Approximately 4 nmol of the transcript **7** from the large-scale transcription reaction was digested with snake venom phosphodiesterase I, calf intestine alkaline phosphatase, and RNase A in 50 mM Tris-HCl buffer (pH 8.5, 50 mM $MgCl_2$, 0.1 mM EDTA) for 15 h at 37°C. The mixture was further treated with RNase T1 for 4 h at 37°C. The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC using Vydac C18 column (0.46 × 25 cm, 5 µm TP silica) at 260 nm and 394 nm. Mobile phase A: 100 mM triethyl ammonium acetate buffer (pH 7.0), mobile phase B: acetonitrile. Flow rate: 1 mL/min. Gradient: 0−7.5% B in 20 min and 7.5−100% in 10 min.

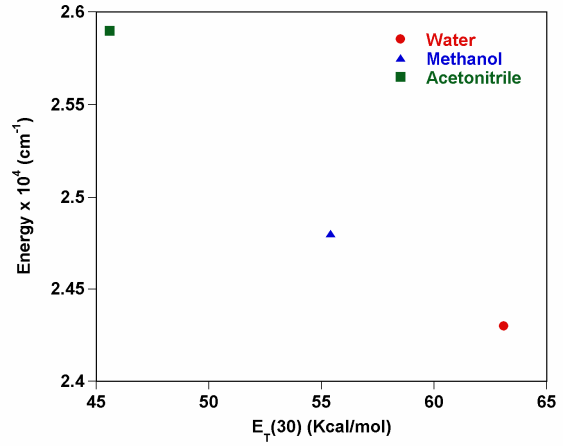


Figure S1. Emission energy plotted against $E_T(30)$, a microscopic solvent polarity scale.

Transcription with different templates: The sequence of the DNA template, particularly near the promoter region significantly controls the efficiency of a

transcription reaction. To evaluate the potential of the enzyme to incorporate the modified uridine at a single position and as well as multiple positions, we designed templates, which would lead to the incorporation of **5** in positions +2 to +9 (Figure S2). Attempts to incorporate the modified UTP analogue closer to the promoter region did not result in the formation of the full length product (Figure S3, lanes 6, 8 and 10). Also, we were unable to effect multiple modifications in the RNA transcript by transcription reaction (Figure S3, lanes 12, 14 and 16). Above results reveal that the triphosphate **5** can be introduced into RNA strand by transcription reaction with moderate efficiency, but only at positions away from the promoter region.

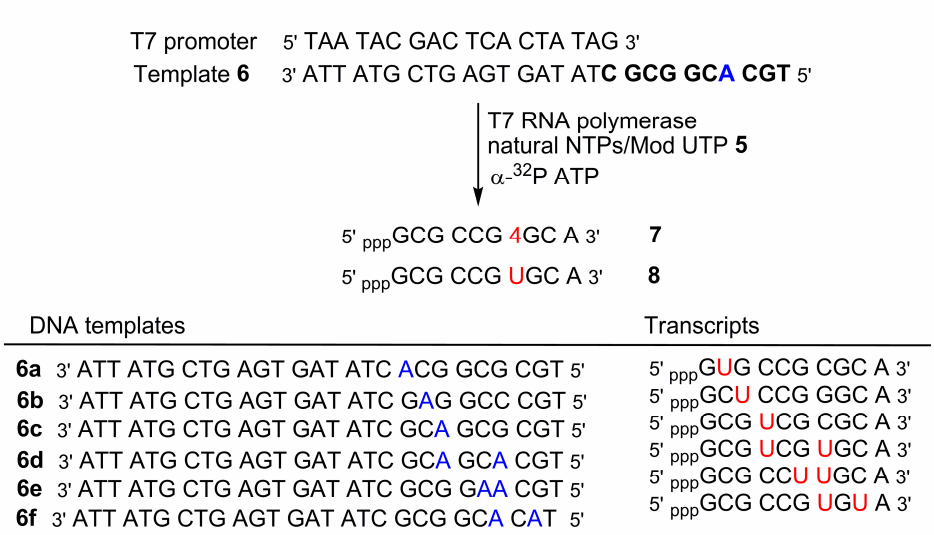


Figure S2. Enzymatic incorporation of thiophene-modified ribonucleoside triphosphate **5** using different templates. DNA templates **6a-6f** annealed to 18mer consensus T7 promoter and corresponding RNA transcripts.

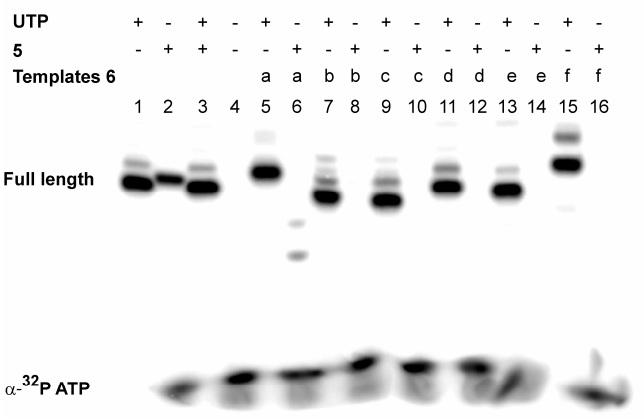


Figure S3. Transcription reaction to study the incorporation efficiency of thiophene-modified triphosphate **5** at various positions in the transcript using different templates.

MALDI-TOF MS measurements: Molecular weight of control and modified RNA transcripts were determined via MALDI-TOF MS. 1 µL of a ~200 µM stock solution of the transcript was combined with 1 µL of 100 mM ammonium citrate buffer (PE Biosystems), 1 µL of a 75 µM DNA standard (17-mer) and 4 µL of saturated 3-hydroxypiccolinic acid. The samples were desalted with an ionexchange resin (PE Biosystems) and spotted onto a gold-coated plate where they were air dried. The resulting spectra were calibrated relative to the +1 and +2 ions of the internal DNA standard, thus the observed oligonucleotides should have a resolution of ±2 mass units.

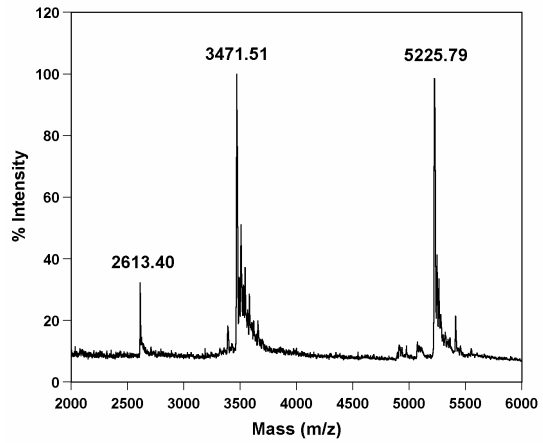


Figure S4. MALDI-TOF MS spectrum of the modified RNA transcript **7** calibrated relative to the +1 and +2 ions of an internal 17-mer DNA standard (m/z: 5225.79 and 2613.40). Calculated mass [M] = 3469.35; observed mass = 3471.51.

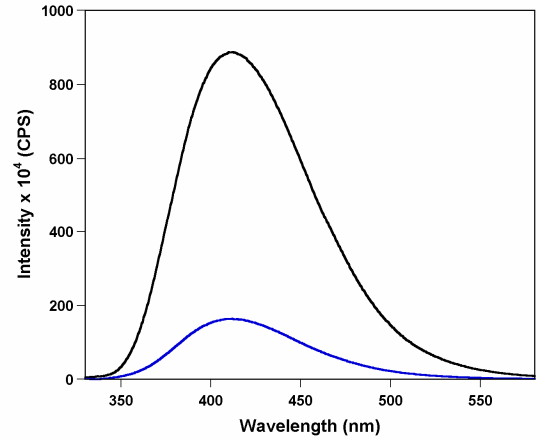


Figure S5. Emission spectra of 1.0 μM solution of nucleoside **4** (black) and RNA transcript **7** (blue) in 20 mM HEPES buffer (100 mM NaCl, 0.5 mM EDTA, pH 7.4). Excitation wavelength was 304 nm, and excitation and emission slit widths were 6 and 8 nm, respectively.

Fluorescence of thiophene-modified duplexes

Transcript RNA **7** was annealed to custom DNA oligonucleotides (**9**−**12**) by heating a 1:1 mixture (12.5 μM) of the oligonucleotides in 20 mM cacodylate buffer (pH 7.0, 500 mM NaCl, 0.5 mM EDTA) at 90°C for 3 min and cooling the samples slowly to room temperature, followed by incubating the solutions in crushed ice. Samples were diluted to give a final concentration of 1 μM in cacodylate buffer. Thiophene-modified duplex constructs (1 μM) were excited at 304 nm with an excitation slit width of 10 nm and emission slit width of 12 nm, and the fluorescence was monitored at the emission maximum, 412 nm. Fluorescence experiments were carried out in a micro fluorescence cell with a path length of 1.0 cm (Hellma GmH & Co KG, Mullenheim, Germany) on a Horiba Jobin Yvon (FluoroMax-3) spectrometer.

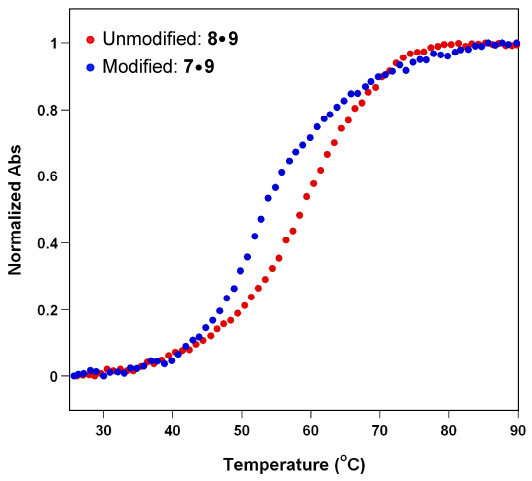


Figure S6. Thermal melting of unmodified (red) and modified (blue) duplex constructs formed by annealing unmodified RNA transcript **8** and thiophene-modified transcript **7** with its complementary DNA strand **9** in 20 mM cacodylate buffer (pH 7, 500 mM NaCl, 0.5 mM EDTA). Duplexes were formed by heating a 1:1 mixture of the oligonucleotides at 90°C for 3 min and cooling the solutions slowly to room temperature. Concentration of each duplex was 1 μM.

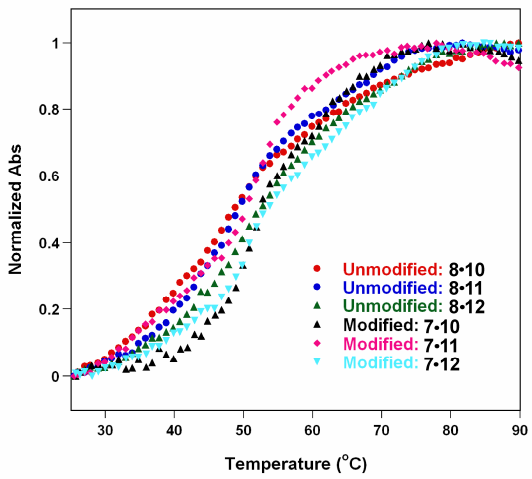


Figure S7. Thermal melting of unmodified and modified duplex constructs in 20 mM cacodylate buffer (pH 7, 500 mM NaCl, 0.5 mM EDTA). Duplexes were formed by heating a 1:1 mixture of the oligonucleotides at 90°C for 3 min and cooling the solutions slowly to room temperature. Concentration of each duplex was 1 μM.

Gel mobility shift experiments: 3'-end ^{32}P-labeled transcript RNA **7** was synthesized by transcription reaction using α-^{32}P ATP. Non-radiolabeled transcript RNA **7** was doped with ^{32}P-labeled RNA **7**, and was annealed to custom DNA oligonucleotides (**9**−**12**) by heating a 1:1 mixture (15 μM) in 20 mM cacodylate buffer (pH 7.0, 500 mM NaCl, 0.5

mM EDTA) at 90°C for 3 min and cooling the samples slowly to room temperature, followed by incubating the solutions in ice. To each sample was added 3 μL of the loading buffer (10 mM Tris-HCl, pH 8, 30% sucrose, 10% glycerol, 0.05% bromophenol blue) and samples (5 μL) were loaded onto an analytical 20% non-denaturing polyacrylamide gel containing 500 mM NaCl. The gel was electrophoresed in TBE buffer (90 mM Tris, 90 mM boric acid and 2 mM EDTA) containing 500 mM NaCl at 125 V. The buffer was recirculated using peristaltic pump. The products on the gel were analyzed using a phosphorimager.

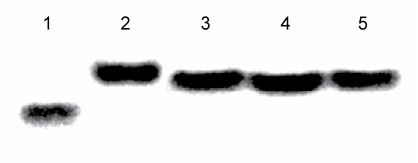


Figure S8. Gel mobility shift analysis of duplexes derived by annealing ^{32}P-labeled transcript RNA **7** to its complementary and mismatched DNA oligonucleotides. Lane 1, ssRNA **7**; lane 2, duplex **7•9**; lanes 3−5, mismatch duplexes **7•11**, **7•12** and **7•10**, respectively.