Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex

SUPPORTING INFORMATION

Wen Zhang[a,b,c,d], Chun Pong Tam[a,b,c,e], Travis Walton[a,b,c,d], Albert C. Fahrenbach[a,b,c,d,f], Gabriel Birrane[g] , Jack W. Szostak[a,b,c,d,e,f,1]

[a]Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, MA 02114; [b]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114; [c]Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA 02114; [d]Department of Genetics, Harvard Medical School, Boston, MA 02115; [e]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; [f]Earth–Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8550, Japan; and [g]Division of Experimental Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215

[1]To whom correspondence should be addressed. Email: szostak@molbio.mgh.harvard.edu.

CONTENTS

1. General Methods.

1a. General considerations. All reagents for the synthesis of 5'-5'-linked dinucleotides were purchased from Sigma Aldrich (St Louis, MO) and used without further purification, except as noted below. Reactions were conducted in oven-dried 4 mL borosilicate glass vials that were fitted with Teflon-lined caps unless otherwise noted. Reagents and materials used for solid-phase RNA polymerization chemistry, including 50 μmol-scale universal controlled-pore glass (CPG) solid support columns, 5′-DMTr-2′-TBDMS-protected RNA phosphoramidites (bz-A-CE, ac-C-CE, ibu-G-CE, and U-CE), acetonitrile, 0.25 M ethylthio-1*H*-tetrazole in acetonitrile (activator solution), 0.02 M iodine in THF/H_2O/pyridine (oxidizing solution), 3% trichloroacetic acid in dichloromethane (deblock solution), cap mix A (THF/acetic anhydride/pyridine 8:1:1), and cap mix B (1-methylimidazole/THF/pyridine 8:1:1) were obtained from Bioautomation (Irving, TX). Reagents for RNA column cleavage, protective group removal and purification, including 28% aqueous ammonium hydroxide, 40% aqueous methylamine, anhydrous dimethyl sulfoxide (DMSO), and triethylamine trihydrofluoride (TEA·3HF), 3 M aqueous sodium acetate, 1-butanol and absolute ethanol, were purchased from Sigma Aldrich. Deuterated solvents were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Preparatory-scale high performance liquid chromatography (HPLC) was carried out on a Varian Prostar 210 HPLC system, equipped with either a preparative-scale Agilent ZORBAX Eclipse-XDB C18 column (21.2x250mm, 7 μm particle size) for reversed-phase chromatography, or with a ThermoFisher-Dionex DNAPac PA100 strong anion exchange column (22x250 mm, 13.5 μm particle size) for anion-exchange chromatography.

1b. NMR data for synthetic dinucleotides. NMR spectra were recorded on a Varian Inova 400 MHz spectrometer (400 MHz for ^{1}H, 100 MHz for ^{13}C, 161 MHz for ^{31}P; Santa Clara, CA). Proton and carbon chemical shifts are reported in parts per million (ppm) values on the δ scale, internally referenced to residual protium in the NMR solvents (Proton NMR: DHO, δ = 4.79 ppm; Carbon NMR: CD_3OD, δ = 49.0 ppm) (1), while proton-decoupled phosphorus chemical shifts were referenced to trimethyl phosphate (D_2O: δ = 3.8 ppm) (2). All NMR spectra were recorded at 25 ^{o}C. Data were reported as

follows: chemical shift, multiplicity (s = singlet, d = doublet, q = quartet, m = multiplet, br = broad), and integration.

1c. RNA oligonucleotides and pGpG synthesis. RNA oligonucleotides and pGpG were synthesized by standard solid-phase phosphoramidite chemistry on a MerMade 6 RNA/DNA oligonucleotide synthesizer (Bioautomation, Irving, TX). Cleavage and elution of the 5´-DMTr-deprotected products from 50 µmol universal CPG-solid support columns were performed by equilibrating and eluting the solid support material a total of 3 times with a 1:1 mixture of ammonium hydroxide and 40% aqueous methylamine (equilibration time: 3 x 10 m; elution volume: 3 x 5 mL for 50 µmol columns). Removal of protecting groups on the nucleobases and phosphates was carried out by heating the basic eluent for 2.5 h at 65 °C; the resultant clear (or off-white) homogeneous mixtures were first concentrated under reduced pressure at 40 ^{o}C for 3 h on a Genevac EZ-2 tabletop speedvac system (Genevac, Stone Ridge, NY), then lyophilized to dryness on a VirTis Sentry 2.0 freeze-drier (SP Scientific, Warminster, PA) at <50 mTorr overnight to afford off-white solid residues. The residues were then resuspended in 2.5 mL of DMSO and 2.5 mL of TEA·3HF, and heated for 2.5 h at 65 °C to remove the TBDMS protecting group on the ribose 2´-hydroxyl group. The mixtures were homogeneous and pale to golden yellow in color. After cooling to room temperature (~30 m), 625 µL of 3 M sodium acetate and 15 mL of 1-butanol were added for RNA precipitation. The precipitates were spun down (4000 rpm, 5 m) and supernatants were removed by decanting. The resulting white solids were washed twice with absolute ethanol; the samples were then dried under high vacuum overnight. Purification of the desired products was carried out by preparative-scale HPLC on a Varian Prostar 210 HPLC system equipped with Agilent ZORBAX Eclipse-XDB C18 column using 25 mM triethylammonium bicarbonate in H_2O (TEAB, pH 7.5) with an increasing gradient of 0 % to 15 % acetonitrile over 30 m. Elution of RNA was monitored by UV absorption at 254 and 280 nm. The desired RNA fractions were collected, pooled and lyophilized to afford a fluffy white powder. The resultant white residues were further purified by preparative-scale strong anion-exchange HPLC on a ThermoFisher-Dionex DNAPac PA100 strong anion exchange column with an increasing gradient of 0 to 100 mM aqueous sodium perchlorate solution over 30 m. The desired RNA fractions were collected, pooled and

lyophilized to afford white solid residues. The residues were washed 3 times with acetone, followed by overnight drying under high vacuum, to afford the desired products (in sodium cation form) as fine white grains.

1d. Locked nucleic acid (LNA)-modified RNA oligonucleotides. The LNA-modified RNA oligonucleotide (LNA in bolded letters) used for crystallographic studies (5′-**mCmCmC**GACUUAAGUCG-3′) was custom-synthesized by Exiqon Inc. (Woburn, MA), with the 5′-dimethoxytrityl (DMT) groups cleaved and samples preliminarily purified by desalting. HPLC purification followed the same procedures as described in section S1c. Concentrations of the aqueous RNA samples were determined by their UV absorption at 260 nm on a Thermo Scientific Nanodrop 2000c Spectrophotometer (Waltham, MA). The theoretical molar extinction coefficients of the RNA strands used herein at 260 nm were provided by Exiqon Inc. (Woburn, MA).

1e. High-resolution mass spectrometry (HRMS) analyses. UHPLC grade (Optima® Grade, Fisher Scientific) reagents and solvents were used to prepare the aqueous buffers and organic solvents for HPLC-TOF-MS analysis, including water, triethylamine, 1,1,1,3,3,3-hexafluoroisopropanol (HFIP), and methanol. High-resolution mass data were obtained for oligonucleotides and synthetic dinucleotides. The analyses were performed on an Agilent 1200 HPLC system coupled to an Agilent 6220 Accurate-Mass time-of-flight mass spectrometer, with a solvent degasser, temperature-controlled auto sampler, column oven, diode-array detector, and a dual electrospray ionization source. Samples were analyzed over a 100 mm XBridge C18 column (1 mm i.d., 3.5 μm particle size, Waters Corporation) using reverse-phase ion-pairing chromatography (3); solvent A was water with 200 mM HFIP, and 1.25 mM TEA at pH 7.0 and solvent B was methanol. For oligonucleotides, solvent B was ramped from 2.5% to 20% over 30 m at a flowrate of 0.1 mL m^{-1} with the column heated to 50 °C. For dinucleotides, isocratic elution with 2% MeOH in solvent A was used. Typically, 100–200 pmole of the analyte was injected for analysis in extended dynamic range in negative ion mode using the following settings: scan rate, 1 spectrum s^{-1}; mass range, 239 m/z – 3200 m/z; drying gas flow, 8 L m^{-1}; drying gas temperature, 325 °C; nebulizer pressure, 30 psig; capillary voltage, 3500 V; fragmentor, 200 V; and skimmer, 65 V. Data analysis was performed using MassHunter Qualitative Analysis (Agilent Technologies). The UV-Vis absorption

and total ionization count (TIC) traces of the samples were monitored and used to gauge purity; all RNA samples used in this study were determined to have a purity of >90%.

1f. NMR Titration Studies. The initial duplex solution contained 1.5 mM RNA duplex, 500 mM sodium chloride, and 10% D_2O; the ligand solution contained the appropriate concentration of GpppG (sodium salt form), as well as the same concentration of the RNA duplex (1.5 mM), D_2O (10%), and sodium cation (500 mM) as that of the starting duplex solution, in order to maintain a constant duplex concentration and ionic strength throughout the titration experiment. A monomer solution containing 30 mM of GpppG was assumed to contain 90 mM of Na^+, and 410 mM of sodium chloride was added to the monomer solution to bring the concentration of Na^+ of the monomer solution to 500 mM. The pH of both duplex and monomer solutions was adjusted to 7.0 (± 0.1) using trace amounts of either aqueous sodium hydroxide or hydrochloric acid. NMR spectra were acquired on a Varian INOVA 400 MHz NMR spectrometer equipped with a broadband PFG (z-gradient) probe. Suppression of the bulk water resonance was achieved using a Watergate pulse sequence (4, 5). Each spectrum was recorded after 128–256 scans, with an optimized delay period on pulse sequence (d1) of 1.0 s and a pulse width (pw) of 15 μs. Initial concentrations of the duplex and monomer solutions were determined by their UV absorption at 260 nm on a Thermo Scientific Nanodrop 2000c Spectrophotometer (Waltham, MA). The theoretical molar extinction coefficients at 260 nm of the RNA strands were calculated with Integrated DNA Technologies' OligoAnalyzer 3.1 (Coralville, IA)(6, 7); the molar extinction coefficients of pGpG, GppG, and GpppG at 260 nm were assumed to be 24160 L mol^{-1} cm^{-1}. The chemicals shifts were referenced externally using a co-axial NMR tube containing a solution of pentafluorobenzaldehyde (δ = 10.285 ppm) in $CDCl_3$.

2. Synthesis of Dinucleotides: Experimental procedures and compound characterizations.

2i. 5′-*O*-Phosphonoguanylyl-(3′→5′)-guanosine (pGpG)

Using standard solid-phase phosphoramidite polymerization chemistry, pGpG was prepared by four 50-µmol-scale solid-phase syntheses (total scale: 0.2 mmol) on a MerMade 6 oligonucleotide synthesizer, in accordance to procedures laid out in section S1c of this Supporting Information.

^{1}H NMR (D_2O)	δ 8.11 (s, 1H), 7.99 (s, 1H), 5.84 (d, *J* = 6.6 Hz, 1H), 5.82 (d, *J* = 5.6 Hz, 1H), 4.7 (dd, *J* = 5.7 & 5.7 Hz), 4.46 (dd, *J* = 3.9 & 4.3 Hz, 1H), 4.41 (br s, 1H), 4.28 (br s, 1H), 4.14 (br s, 2H), 3.92 (br s, 2H). *Protons on the 2'-position of the two riboses (~ δ 4.84 – 4.80) were poorly resolved from the HOD signal at δ 4.79 ppm.*
^{13}C NMR (D_2O)	δ 162.20, 162.05, 156.89, 156.78, 152.80, 152.65, 138.02, 137.88, 117.48, 117.40, 87.71, 86.93, 84.88 (dd, *J* = 9 & 3 Hz), 84.65 (d, *J* = 9 Hz), 76.33 (d, *J* = 5 Hz), 74.51, 74.03 (d, *J* = 5 Hz), 71.38, 66.12 (d, *J* = 5 Hz), 64.5 (d, *J* = 4 Hz)
^{31}P NMR (D_2O)	δ 0.34, –4.54
HRMS (*m/z*)	Calc'd for $C_{20}H_{25}N_{10}O_{15}P_2$ $[M–H]^–$: 707.0982; Found: 707.1016

2ii. P^1,P^2-Diguanosine-5'-pyrophosphate (GppG)

The synthetic protocol was adapted from the report of Tanaka et al. (8). 2-Chloro-1,3-dimethylimidazolinium chloride (DMC, 340 mg, 2 mmol, 2 equiv.; *very hygroscopic*), imidazole (272.3 mg, 4 mmol, 4 equiv.) and guanosine 5'-monophosphate disodium salt (GMP•2Na^+, 407.2 mg, 1 mmol, 1 equiv.) were charged into an oven-dried 4 mL borosilicate glass vial equipped with a stir bar and a Teflon-lined cap. D_2O (2 mL, 0.5 M) was added as a solvent. The mixture was stirred at 40 ^{o}C for 1 h, followed by addition of another 1 equiv. of GMP•2Na^+ (407.18 mg). After further stirring at 40 ^{o}C overnight, the reaction mixture was diluted 20-fold with water, followed by preparative-scale HPLC purification on a Varian Prostar 210 HPLC system equipped with Agilent ZORBAX

Eclipse-XDB C18 column using 25 mM triethylammonium bicarbonate in H_2O (pH 7.5) with an increasing gradient of 0 % to 15 % acetonitrile over 30 m. Elution of RNA was monitored by UV absorption at 254 and 280 nm. The desired RNA fractions were collected, pooled and lyophilized to afford a fluffy white powder. The resultant white residues were again purified with preparative-scale strong anion-exchange HPLC with a ThermoFisher-Dionex DNAPac PA100 strong anion exchange column with an increasing gradient of 0 to 100 mM aqueous sodium perchlorate over 30 m. The desired fractions were collected, pooled and lyophilized at <50 mTorr to afford white solid residues. The residues were washed with acetone for 3 times, followed by overnight drying under high vacuum, to afford the desired products (in sodium cation form) as fine white grains.

^{1}H NMR (D_2O)	δ 7.94 (s, 2H), 5.78 (d, *J* = 5.2 Hz, 2H), 4.62 (dd, *J* = 5.1 & 5.1 Hz, 2H), 4.44 (dd, *J* = 4.4 & 4.6 Hz, 2H), 4.31–4.26 (m, 4H), 4.22–4.17 (m, 2H)
^{13}CNMR (D_2O)	δ 159.5, 154.6, 152.1, 138.1, 116.9, 88.3, 84.1 (t, *J* = 4.5 Hz), 74.9, 70.9, 65.9
^{31}P NMR (D_2O)	δ –10.50
HRMS (*m/z*)	Calc'd for $C_{20}H_{25}N_{10}O_{15}P_2$ $[M–H]^-$: 707.0982; Found: 707.1013

2iii. P^1,P^3-Diguanosine-5'-triphosphate (GpppG)

The synthetic procedures leading to GpppG were highly analogous to that of GppG, except that guanosine 5'-diphosphate (sodium salt form) was added in lieu of GMP•2Na^+ following the initial one-hour DMC-mediated GMP activation.

^{1}H NMR (D_2O)	δ 8.05 (s, 2H), 5.88 (d, *J* = 4.8 Hz, 2H), 4.67 (dd, *J* = 4.9 & 4.9 Hz, 2H), 4.5 (dd, *J* = 4.6 & 4.7 Hz, 2H), 4.36 (br, 2H), 4.33 (m, 4H)
^{13}CNMR (D_2O)	δ 159.4, 154.6, 152.1, 138.1, 116.8, 88.2, 84.0 (d, *J* = 9 Hz), 74.9, 70.7, 65.7
^{31}P NMR (D_2O)	δ –10.45 (d, *J* = 9.7 Hz), –21.76 (t, *J* = 17.7 Hz)
HRMS	Calc'd for $C_{22}H_{29}N_7O_{10}P$ $[M–H]^-$: 787.0645; Found: 787.0675

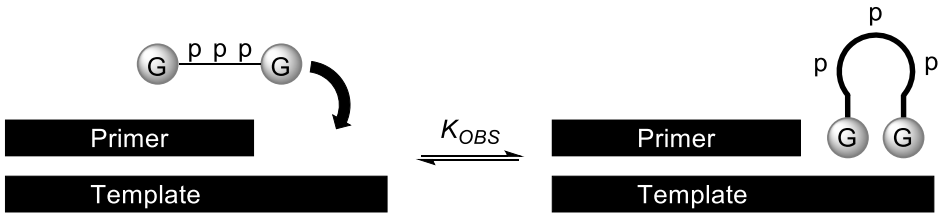
2iv. 1,3-di-(guanosine-5'-phosphoryl)-2-aminoimidazolium (Gp-NH_2Im-pG)

The synthesis of the di-guanosine intermediate began by first synthesizing two monomers, guanosine-5'-phosphoryl-(2-aminoimidazole) (2-AmImpG) and guanosine-5'-phosphoryl-(1-hydroxy-7-azabenzotriazole) (GMP-OAt). Aqueous mixtures of GMP•$2H^+$ (100 mg, 0.275 mmol, 1 equiv.) with either 2-aminoimidazole (dark brown oil, 91 mg, 1.1 mmol, 4 equiv.) or 1-hydroxy-7-azabenzotriazole (200 mg, 1.47 mmol, 5.3 equiv.), and triethylamine (200 µL, 1.4 mmol, 4 equiv., ρ = 0.726 g mL^{-1}) were first prepared, vortexed and sonicated until complete homogenization, then flash frozen in liquid nitrogen and lyophilized over 5 days at < 50 mTorr. Once dry, the solids were separately resuspended in DMSO (20 mL, 14 mM) and triethylamine (300 µL, 2.15 mmol, 7.8 equiv., ρ = 0.726 g mL^{-1}). 2,2'-dipyridyl disulfide (1.2 g, 5.5 mmol, 20 equiv.) and triphenylphosphine (1.2 g, 4.6 mmol, 17 equiv.) were added and the reactions were left stirring overnight (~ 12 hrs). Extra 2,2'-dipyridyldisulfide (0.62 g, 2.8 mmol, 10.2 equiv.) and triphenylphosphine (0.6 g, 2.3 mmol, 8.4 equiv.) were added to the reactions. Four hours later, both reactions were separately precipitated in precipitation solutions containing 120 mL acetone, 60 mL diethyl ether, and 4.5 g of sodium perchlorate. Pellets were washed twice with acetone, followed by house vacuum-drying overnight. Afterwards, one half of the crude monomer GMP-OAt and one half of the crude 2-AmImpG were mixed together in 5 mL H_2O. This mixture was incubated for 70 m at room temperature and became viscous. The mixture was then purified by reverse phase flash chromatography, by direct loading onto a 30 g C18aq column on a Combiflash Rf-200 from Teledyne Isco. The products were eluted over 15 column volumes, with a gradient of 20 mM TEAB (pH = 7.5) over 0 to 20% acetonitrile at a flow rate of 20 mL/m. The fraction containing the di-guanosine intermediate was immediately flash frozen and lyophilized at -20°C on a VirTis AdVantage Plus EL-85 lyophilizer from SP Scientific. Excess triethylamine from the previous purification was removed by reverse phase chromatography again, using the same procedure, except that the aqueous solvent was water instead of TEAB. The fraction containing the intermediate was then flash frozen and lyophilized at -20°C.

^{1}H NMR (D_2O)	δ 7.68 (s, 2H), 6.56 (dd, *J* = 2.6, 1.8 Hz, 2H), 5.59 (d, *J* = 5.2 Hz, 2H), 4.51 (t, *J* = 5.2 Hz, 2H), 4.26 (dd, *J* = 5.0, 4.5 Hz, 2H), 3.97–3.95 (m, 4H), 3.88-3.83 (m, 2H)
^{13}C NMR (D_2O)	δ 159.4, 154.3, 152.2, 150.7 (weak, t, *J* = 7 Hz),138.0, 116.8, 116.4–116.6 (m), 87.9, 83.2 (d, *J* = 8.3 Hz), 73.9, 70.4, 66.5 (d, *J* = 6.4 Hz)
^{31}P NMR (D_2O)	δ –9.04
HRMS (*m/z*)	Calc'd for $C_{23}H_{28}N_{13}O_{14}P_2$ $[M–H]^–$: 772.1359; Found: 772.1371

3. Affinity of GpppG for the RNA duplex P/T_{2C}

3a. Derivation of the NMR binding isotherm.(9)



Scheme S1. The proposed binding scheme between **P/T_{2C}** duplex with GpppG.

We have previously reported the use of the **P/T_{2C}** duplex (Sequence: **P**: 5'–CUCAAUG–3'; **T_{2C}**: 5'–CCCAUUGAG–3') to monitor the binding affinity of two consecutive GMP molecules onto the RNA duplex (10). Upon GMP binding, the imino proton signal of G7 (the 3'-terminal nucleotide of the primer, flanking the ligand binding sites) was observed to shift upfield, without observation of new imino proton signals. This observation is in line with our hypothesis that GMP binding with RNA duplexes is in fast exchange, with the on- and off-rates of ligand binding being faster than the NMR time scale. When these conditions are met, and when the ligand-duplex binding stoichiometry is expected to be 1:1, the following single-site binding isotherm is commonly used to numerically approximate the binding affinity (*K*):

$$\Delta\delta = \frac{\Delta\delta_{tot} K[M]}{1 + K[M]}$$

with Δδ being the measured chemical shift change induced by monomer binding, $\Delta\delta_{tot}$ being the theoretical total chemical shift change when all duplex binding sites are fully

saturated, and [M] being the concentration of *free*, *unbound* monomer (*not* the concentration of monomer added to the duplex solution). Since we expected the binding of GpppG onto **P/T$_{2C}$** to be tight, we hypothesized that a substantial portion of the added monomer would be bound to the duplex, with only a small fraction of the added monomer being left in the unbound state. We therefore need to derive an equation which will explicitly solve the concentration of unbound monomer at all points of the titration.

We begin the derivation by realizing that the binding of GpppG with the **P/T$_{2C}$** duplex can be represented by the following equilibrium:

$$[D] + [M] \underset{}{\overset{K}{\rightleftharpoons}} [DM]$$

where D represents **P/T$_{2C}$** duplex, M represents GpppG, and DM represents GpppG-bound RNA duplex. *K* is equal to:

$$K = \frac{[DM]}{[D] \cdot [M]}$$

At any point of the titration, the total concentration of duplex (D_{tot}) is equal to the sum of the concentration of GpppG-bound RNA duplex (DM), and free RNA duplex (D). The same applies to monomer as well. Hence:

$$[D] + [DM] = [D_{tot}]\,;\; [M] + [DM] = [M_{tot}]$$

Substituting these equations into the equilibrium expression, we get:

$$K = \frac{[DM]}{(D_{tot} - [DM]) \cdot (M_{tot} - [DM])}$$

Rearrangement of this expression in terms of [DM] gives a quadratic equation:

$$[DM]^2 - \left(M_{tot} + D_{tot} + \frac{1}{K}\right)[DM] + M_{tot}D_{tot} = 0$$

and the roots (solution) of this quadratic expression can be expressed as the following:

$$[DM] = \frac{\left(M_{tot} + D_{tot} + \frac{1}{K}\right) \pm \left[\left(M_{tot} + D_{tot} + \frac{1}{K}\right)^2 - 4M_{tot}D_{tot}\right]^{\frac{1}{2}}}{2}$$

As K approaches positive infinity, the two roots bifurcate into M_{tot} (larger root) and D_{tot} (smaller root); however, since $M_{tot} > D_{tot}$ at the end of the titration, [DM] must be limited by the total duplex concentration (D_{tot}). Hence [DM] cannot be equal to M_{tot}, and the larger root is rejected. Hence,

$$[DM] = \frac{\left(M_{tot} + D_{tot} + \frac{1}{K}\right) - \left[\left(M_{tot} + D_{tot} + \frac{1}{K}\right)^2 - 4M_{tot}D_{tot}\right]^{\frac{1}{2}}}{2}$$

Subsequently, at any point of the NMR titration, the concentration of free, unbound monomer is equal to

$$[M] = [M_{tot}] - [DM]$$

$$[M] = \frac{M_{tot} - D_{tot} - \frac{1}{K} + \left[\left(M_{tot} + D_{tot} + \frac{1}{K}\right)^2 - 4M_{tot}D_{tot}\right]^{\frac{1}{2}}}{2}$$

This equation is substituted back into the canonical NMR binding isotherm:

$$\Delta\delta = \frac{\Delta\delta_{tot}K[M]}{1 + K[M]}$$

and we get the crude NMR binding isotherm which is then used to fit the NMR data for numerical approximation of the GpppG–RNA duplex binding affinity:

$$\Delta\delta = \frac{\frac{1}{2}\Delta\delta_{tot}K\left\{M_{tot} - D_{tot} - \frac{1}{K} + \left[\left(M_{tot} + D_{tot} + \frac{1}{K}\right)^2 - 4M_{tot}D_{tot}\right]^{\frac{1}{2}}\right\}}{1 + \frac{1}{2}K\left\{M_{tot} - D_{tot} - \frac{1}{K} + \left[\left(M_{tot} + D_{tot} + \frac{1}{K}\right)^2 - 4M_{tot}D_{tot}\right]^{\frac{1}{2}}\right\}}$$

$$\Delta\delta = \frac{\Delta\delta_{tot}K\left\{M_{tot} - D_{tot} - \frac{1}{K} + \left[\left(M_{tot} + D_{tot} + \frac{1}{K}\right)^2 - 4M_{tot}D_{tot}\right]^{\frac{1}{2}}\right\}}{2 + K\left\{M_{tot} - D_{tot} - \frac{1}{K} + \left[\left(M_{tot} + D_{tot} + \frac{1}{K}\right)^2 - 4M_{tot}D_{tot}\right]^{\frac{1}{2}}\right\}}$$

The true total monomer concentration (M_{tot}) throughout the titration may deviate from the theoretical value, due to factors like handling error in monomer addition, or inaccuracies in the approximation of the true concentration of the ligand in the monomer solution. To account for these possible deviations, the stoichiometric factor *n* is introduced to the D_{tot} term, yielding the final form:

$$\Delta\delta = \frac{\Delta\delta_{tot}K\left\{M_{tot} - nD_{tot} - \frac{1}{K} + \left[\left(M_{tot} + nD_{tot} + \frac{1}{K}\right)^2 - 4nM_{tot}D_{tot}\right]^{\frac{1}{2}}\right\}}{2 + K\left\{M_{tot} - nD_{tot} - \frac{1}{K} + \left[\left(M_{tot} + nD_{tot} + \frac{1}{K}\right)^2 - 4nM_{tot}D_{tot}\right]^{\frac{1}{2}}\right\}}$$

3b. Titration of GpppG into the P/T$_{2C}$ duplex: NMR studies.

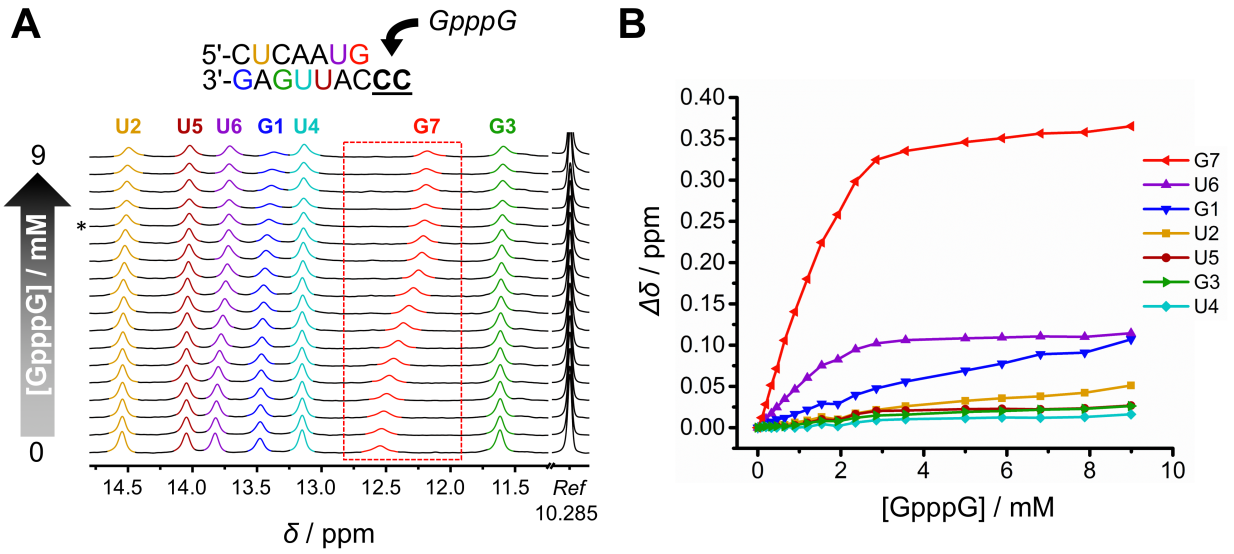


Figure S1. Titration of the **P/T$_{2C}$** duplex with GpppG while monitoring the 10 to 15 ppm region of the ^{1}H NMR spectra.(11, 12) **A.** The component spectra for the GpppG–**P/T$_{2C}$** titration were stacked for convenient visualization. The 5'-CC overhang of **P/T$_{2C}$** engages in binding with GpppG. All spectra were recorded at 12 ^{o}C in the presence of 500 mM Na^{+} and 10% D_2O at pH 7 with a duplex concentration of 1.5 mM. The signal at 10.285 ppm is that of the pentafluorobenzaldehyde reference. **B.** Using the spectral data shown in figure **S1A**, the change in chemical shifts of all **P/T$_{2C}$** imino protons were plotted against the concentration of GpppG. The imino proton signal of G7 (red) shifted the most, followed by the signals of U6 (purple), G1 (blue), and U2 (mustard yellow). The signals of internal imino protons U5, G3, and U4 (crimson, green, and cyan, respectively) shifted only minimally.

To assess the affinity of GpppG with cognate RNA duplexes, we used a previously-reported **P/T$_{2C}$** duplex(10) with a 5'-CC overhang. Under the conditions of the NMR titration experiments (500 mM Na^{+}, pH 7, 12 ^{o}C), the primer and template strands are known to be stably annealed to form an A-form duplex. Additionally, the imino protons of the **P/T$_{2C}$** duplex are clearly defined and baseline-resolved. The assignment

of proton resonances (**figure S1A**) are deduced from data of previously-described variable-temperature proton and two-dimensional ^{1}H-^{1}H NOESY experiments(10, 13).

We carried out a titration from 0 to 9 mM GpppG (0 to 6 equivalents) into 1.5 mM of **P/T$_{2C}$** solution, and the change in chemical shift of the G7 imino proton was fitted to a modified single-site binding isotherm. As the concentration of GpppG increased, G7 shifted upfield remarkably (~0.35 ppm, red trace, **figure S1B**) and in a hyperbolic fashion, which is in line with our previous observations that purine nucleotide ligands tend to shift the G7 resonance upfield (13). The imino proton resonance of U6 also undergoes an upfield and hyperbolic shift, although to a lesser degree (~0.10 ppm, purple trace, **figure S1B**). On the other hand, both G1 and U2 imino protons (blue and mustard yellow traces, **figure S1B**) undergo broadening and upfield shifting as GpppG is titrated into the duplex solution, although these changes are much smaller than those seen for the G7 imino proton. Since both G1 and U2 imino protons are on the blunt-end terminus of **P/T$_{2C}$** with no GpppG binding sites on the template, we hypothesize that these experimental observations likely arise from side-on, non-specific association of GpppG onto the blunt-end terminus of **P/T$_{2C}$**, and is likely unrelated to the Watson Crick based GpppG–**P/T$_{2C}$** binding of interest. Finally, the three "internal" imino protons (G3, U4, and U5; green, cyan, and crimson traces, **figure S1B**) only shifted upfield minimally (< 0.05 ppm), showing that the effect of GpppG binding does not propagate beyond two base pairs.

It is interesting to note that, as the concentration of GpppG in the duplex solution is increased beyond 6 mM (asterisked spectrum, **figure S1A**), all imino proton resonances begin to broaden into the baseline. We are reluctant to speculate on the factors or nonidealities that give rise to these observations; however, we suspect that when GpppG is in excess of available duplex binding sites, and when temperature is lowered, the nonspecific interactions between GpppG and surrounding **P/T$_{2C}$** duplexes could be significant enough to be able to bridge multiple RNA duplexes noncovalently, to form a relatively large RNA-GpppG complex that undergoes slow tumbling on the NMR timescale.

4. K_M measurement for the 2-aminoimidazolium-bridged diguanosine dinucleotide (Gp-NH_2Im-pG)

4a. Primer extension reaction.

Nonenzymatic template-directed RNA polymerization was monitored through primer extension assays. For accurate comparison between the RNA•GpppG and RNA•Gp-NH_2Im-pG, we used the same primer-template duplex as the one used in the NMR binding assay (Sequences: primer, 5′-FAM-CUCAAUG-3′; template, 5′-CCCAUUGAG-3′). The final concentrations of the primer extension reaction were 2 μM primer, 3 μM template, 200 mM Tris pH 8, 100 mM $MgCl_2$, and 0.125–2 mM of the Gp-NH_2Im-pG intermediate, adjusted for purity. After adding Gp-NH_2Im-pG to the reaction mixture to initiate primer extension, 1 μL reaction aliquots were removed at 1, 2, and 3 m and quenched in 7 μL of 8 M urea, 100 mM Tris-Cl, 100 mM boric acid, and 75 mM EDTA. Reaction aliquots were flash frozen on dry ice, and kept under –80 ˚C until analysis by 20% denaturing polyacrylamide gel electrophoresis (National Diagnostics, Atlanta, GA). Gels were imaged on an Amersham Typhoon scanner from General Electric and quantified using the ImageQuantTL software (Little Chalfont, United Kingdom). Data were analyzed assuming pseudo-first order kinetics to determine the initial rate constant, k_{obs}.

4b. Michaelis-Menten analysis of RNA:Gp-NH_2Im-pG binding

To compare the RNA binding affinity of GpppG with that of Gp-NH_2Im-pG, we sought to determine the K_M of this intermediate in primer extension reactions. We were able to obtain relatively pure Gp-NH_2Im-pG through the synthetic route laid out in Section S2, (page S8) *via* two rounds of purification, with ≥ 80% of the purified material in the form of imidazolium-bridged dinucleotide, and the remainder as the monomer (2-AmImpG). In addition, these fractions contained 1ed–2 equivalents of triethylamine. For our assays, we assumed that the RNA binding affinity of the intermediate would be greater than that of the monomer, and that the monomer would contribute insignificantly to the polymerization rate under these conditions.

For the primer extension assay, we used the same primer and template RNA sequences as that in the affinity measurement made by NMR. We observed that the

buffer and pH conditions of the NMR studies (pH 7, no Mg^{2+}) were not compatible with our K_M determination because of the slow polymerization rate of primer extension. In addition, we observed that solutions containing even 100 mM NaCl resulted in precipitation of the intermediate and poor kinetic analysis of primer extension. Therefore, we used 200 mM Tris pH 8 and 100 mM $MgCl_2$ for our primer extension studies. We determined the initial rate of primer extension for 5 concentrations of the Gp-NH_2Im-pG and measured K_M = 619 ± 60 µM (**figure S2**). This value is 3x greater than the K_d = 175 µM of the GpppG analog by NMR, but is 27x less than the K_d = 17 mM of GMP. This suggests that the binding of the intermediate is better approximated by GpppG than GMP. The discrepancy between the K_M of the intermediate and the K_d of GpppG may be due to a variety of factors. For instance, the buffer conditions and pH of the two experimental systems differ. In addition, the relationship between K_M and K_d may not be directly comparable. Therefore, we believe that these values are in rough agreement and support our proposal that GpppG is an analogue of the di-guanosine intermediate.

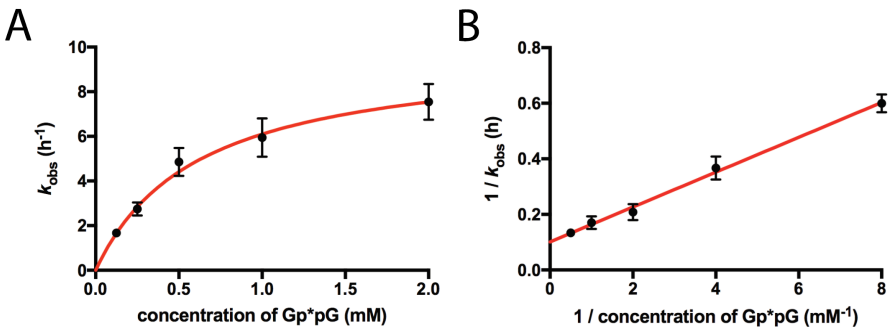


Figure S2. Determination of the K_M of Gp-NH_2Im-pG dinucleotide by the primer extension assay. (A) Michaelis-Menten plot of the k_{obs} of primer extension versus the concentration of the di-guanosine intermediate. Black circles represent average experimental values in triplicate and error bars indicate ± 1 S.D. The red line is our fit of the experimental data using K_M = 0.619 mM and the maximum k_{obs} = 9.87 h^{-1}. (B) Lineweaver-Burke plot of the experimental data from part A. Black circles represent experimental data with error bars ± 1 S.D. The red line is the linear regression of this data.

5. X-ray crystallography.

5a. Crystal preparation. The Nuc-Pro High Throughput Screen Kit (Jena Bioscience), Natrix High Throughput Kit and Index High Throughput Kit (Hampton Research, Aliso Viejo, CA) were used for screening crystallization conditions by the sitting drop vapor diffusion method. Solutions containing the RNA sample (0.5 mM) and different dinucleotide ligands (10 mM) were heated to 90 °C for 2 m, then cooled slowly to room temperature. All of the crystals grew at 18 °C, and the mother liquor containing 50% glycerol was used as a cryoprotectant during crystal mounting. All data collection was taken under a stream of nitrogen at 99 K. The data sets were collected at the SIBYLS beamline 821 and 822 at the Advanced Light Source, Lawrence Berkeley National Laboratory. The distances between the detector and the crystal were set to 200 mm and the collecting wavelength was set to 0.997 Å. The crystals were exposed for 1 second per image with one degree oscillations, and 180 images were taken for each data set. The optimized crystallization conditions for the RNA-monomer complexes are listed below.

Table S1. Optimized conditions for crystallization of RNA-dinucleotide complexes

	Optimized crystallization conditions
RNA–GpppG	0.05 M Magnesium chloride, 0.1 M Imidazole pH 6.5, 1.0 M Sodium acetate trihydrate
RNA–GppG	0.05 M Magnesium chloride, 0.2 M Ammonium citrate tribasic pH 7.0, 20% w/v Polyethylene glycol 3,350
RNA–pGpG	0.05 M Magnesium chloride, 0.1 M HEPES sodium pH 7.5, 2% v/v Polyethylene glycol 400, 2.0 M Ammonium sulfate
RNA–GppppG	0.05 M Magnesium chloride, 1.2 M Lithium sulfate, 50 mM MES pH 6.5, 2 mM Cobalt (II) chloride

5b. Data collection and structure refinement. The data were processed using HKL2000 and DENZO/SCALEPACK. All of the structures were solved by molecular replacement, using structure of 5DHC as search model. All four structures were refined using Refmac. The refinement protocol included simulated annealing refinement, restrained B-factor refinement, and bulk solvent correction. During refinement, the

topologies and parameters for locked nucleic acids (LCC) and for the ligands GpppG (GP3), GppG (GP2), GppppG (GP4) were constructed and applied. After several cycles of refinement, a number of highly ordered water molecules and magnesium ions were added. The 4 crystal structures were determined to resolutions of 1.9 Å, 1.5 Å, 2.6 Å and 2.1 Å, respectively (PDB ID: 5UEE, 5UED, 5UEG, 5UEF). Data collection, phasing, and refinement statistics of the determined structures are listed in Tables S2 and S3.

Table S2. Data collection statistics.

Structure	RNA–GpppG	RNA–GppG	RNA–GppppG	RNA–pGpG
Space group	$P3_121$	$P3_121$	P3	$P3_121$
Unit cell parameters	46.96, 46.96, 83.10	43.91, 43.91, 85.41	48.42, 48.42, 81.96	43.63, 43.63, 84.05
(Å, ^{o})	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range, Å	50-1.90	50-1.50	50-2.60	50-2.10
(last shell)	(1.97-1.90)	(1.55-1.50)	(2.69-2.60)	(2.18-2.10)
Unique reflections	8742	27933	12517	10461
Completeness, %	99.4 (94.4)	95.1 (100)	96.6 (74.8)	99.8 (99.6)
R_{merge}, %	6.3 (46.4)	7.9 (35.1)	10.0 (55.4)	7.4 (46.9)
$<I/\sigma(I)>$	35.3 (1.9)	28.4 (5.5)	15.1 (1.3)	30.9 (2.8)
Redundancy	9.1 (4.5)	9.8 (7.6)	5.1 (3.6)	9.6 (7.0)

Table S3. Structure refinement statistics.

Structure	RNA–GpppG	RNA–GppG	RNA–GppppG	RNA–pGpG
PDB code	5UEE	5UED	5UEG	5UEF
RNA duplex per asymmetric unit	1	1	2	1
Resolution range, Å	83.10-1.90	85.41-1.50	81.96-2.60	84.05-2.10
R_{work}, %	23.20	21.03	20.76	23.97
R_{free}, %	27.47	23.22	30.15	30.32
Number of reflections	8299	14252	6081	5391
Bond length R.M.S. (Å)	0.016	0.016	0.017	0.014
Bond angle R.M.S.	1.854	1.843	2.037	1.589
Average B-factors, (Å^2)	32.42	18.04	84.06	46.67

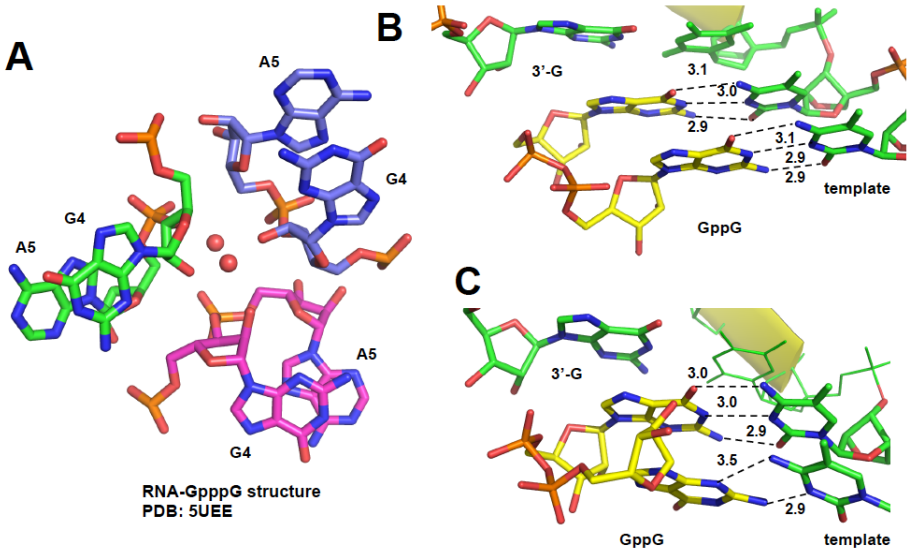


Figure S3. (A) Local structure of RNA-GpppG complex. Two of the water molecules observed to bridge the three neighboring RNA duplexes interact by hydrogen bonding with 2'-hydroxyls of G4s and the G4-G5 phosphodiester phosphates. (B) and (C) Local structures of the RNA-GppG complex. At one end, GppG forms two Watson-Crick base pairs with the template, and at the other end, one Watson-Crick base pair and one noncanocial base pair are observed.

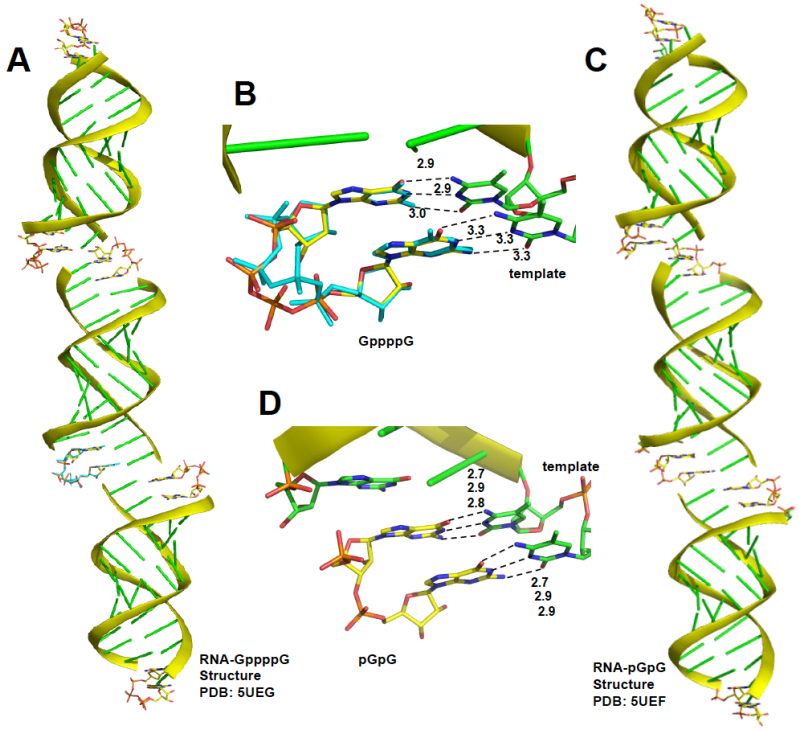


Figure S4. (A) The RNA-GppppG complexes slip-stack to crystallize. (B) The GppppG ligand forms two Watson-Crick base pairs with the template. (C) The RNA-pGpG complexes slip-stack to crystallize. (D) The pGpG dimer forms two Watson-Crick base pairs with the template.

6. References.

1. Gottlieb HE, Kotlyar V, Nudelman A (1997) NMR chemical shifts of common laboratory solvents as trace impurities. *J Org Chem* 62(21):7512-7515.
2. Streck R, Barnes AJ (1999) Solvent effects on infrared, 13C and 31P NMR spectra of trimethyl phosphate: Part 1. Single solvent systems. *Spectrochim Acta A* 55(5):1049-1057.
3. Apffel A, Chakel JA, Fischer S, Lichtenwalter K, Hancock WS (1997) Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry. *Anal Chem* 69(7):1320-1325.
4. Piotto M, Saudek V, Sklenár V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2(6):661-665.
5. Lippens G, Dhalluin C, Wieruszeski JM (1995) Use of a water flip-back pulse in the homonuclear NOESY experiment. *J Biomol NMR* 5(3):327-331.
6. Cantor CR, Warshaw MM, Shapiro H (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* 9(9):1059-1077.
7. Cavaluzzi MJ, Borer PN (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res* 32(1):e13.
8. Tanaka H, Yoshimura Y, Jørgensen MR, Cuesta-Seijo JA, Hindsgaul O (2012) A simple synthesis of sugar nucleoside diphosphates by chemical coupling in water. *Angew Chem Int Ed Engl* 51(46):11531-11534.
9. Anslyn EV, Dougherty DA (2006) *Modern Physical Organic Chemistry* (University Science Books, Sausalito, CA).
10. Tam CP*, et al.* (2017) Downstream Oligonucleotides Strongly Enhance the Affinity of GMP to RNA Primer-Template Complexes. *J Am Chem Soc* 139(2):571-574.
11. Conte MR, Conn GL, Brown T, Lane AN (1996) Hydration of the RNA duplex r(CGCAAAUUUGCG)$_2$ determined by NMR. *Nucleic Acids Res* 24(19):3693-3699.
12. Lane AN, Jenkins TC, Frenkiel TA (1997) Hydration and solution structure of d(CGCAAATTTGCG)$_2$ and its complex with propamidine from NMR and molecular modelling. *Biochim Biophys Acta* 1350(2):205-220.
13. Izgu EC*, et al.* (2015) Uncovering the thermodynamics of monomer binding for RNA replication. *J Am Chem Soc* 137(19):6373-6382.