Supporting Information for

Synthesis of a nonhydrolyzable nucleotide phosphoroimidazolide analogue that catalyzes nonenzymatic RNA primer extension

Chun Pong Tam†‡, Lijun Zhou†§, Albert C. Fahrenbach†§⊥, Wen Zhang†§, Travis Walton†¶, and Jack W. Szostak*†‡§⊥

† Howard Hughes Medical Institute, Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA.
‡ Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA
§ Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA
⊥ Earth Life Science Institute, Tokyo Institute of Technology, 2-12-1-IE-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan
¶ Program of Biological and Biomedical Sciences, Harvard Medical School, 25 Shattuck Street, Gordon Hall, Boston, Massachusetts 02115, USA.
* Email: szostak@molbio.mgh.harvard.edu
‡ These authors contributed equally to this work.

CONTENTS

1. General Methods S2
2. Synthesis of ICG and 2-MelmpddC S5
3. pKₐ Determination of Nonhydrolyzable Analogues S13
4. Isothermal Titration Calorimetry S17
5. Nonenzymatic Primer Extension S23
6. CplCG: Detection and Structure Elucidation S33
7. References S64
1. General Methods

1.1 Reagents

| Reagents / Starting Material | Purpose | Supplier |
|------------------------------|---------|----------|
| • 50 μmole universal controlled-pore glass (CPG) solid support columns | Solid-phase RNA synthesis | Bioautomation (Irving, TX) |
| • 5’-DMTr-2’-TBDMS-protected RNA phosphoramidites (bz-A-CE, ac-C-CE, ibu-G-CE, and U-CE) |  | EMD Serono (Darmstadt, Germany) |
| • Acetonitrile, Ultra-low-water |  |  |
| • 3% trichloroacetic acid in dichloromethane (deblock) |  |  |
| • 0.25 M ethylthio-1H-tetrazole in acetonitrile (activator) |  |  |
| • 0.02 M iodine in THF/H₂O/pyridine (oxidizer) |  |  |
| • 8:1:1 THF/acetic anhydride/pyridine (cap mix A) |  |  |
| • 8:1:1 N-methylimidazole/THF/pyridine (cap mix B) |  |  |
| • 28% aqueous ammonium hydroxide | RNA elution and deprotection | Sigma Aldrich (St. Louis, MO) |
| • 40% aqueous methylamine |  |  |
| • dimethyl sulfoxide, 1-butanol, ethanol |  |  |
| • triethylamine trihydrofluoride |  |  |
| • 3 M aqueous sodium acetate |  |  |
| • Deuterium-labelled solvents | NMR analysis | Cambridge Isotope Laboratories (Tewksbury, MA) |
| • 32% (w/v) Deuterium chloride in D₂O |  |  |
| • 50% (w/w) Sodium deuteroxide in D₂O |  |  |
| • Triethylamine, acetonitrile, water, methanol, 1,1,1,3,3,3-hexafluoropropanol (All in Optima® Grade) | HPLC-TOF analysis | Fisher Scientific (Waltham, MA) |
| • Custom-synthesized RNA oligonucleotides | Primer extension assays | Integrated DNA Technologies (Coralville, IA) |
| • 1 M aqueous magnesium chloride |  | Ambion (Grand Island, NY) |
| • Na•MES (pH 5.5, 6.3) and Na•HEPES (pH 7.0, 8.0) |  | Hampton Research (Aliso Viejo, CA) |
| • Starting material and reagents | Synthesis of ICG and 2-MelmpN | Sigma Aldrich (St. Louis, MO) |
| • SureSeal™ anhydrous solvents |  | Chem-Impex International (Wood Dale, IL) |
| • Glass-backed precoated high-performance silica plates with blue fluorescent indicators |  | EMD Millipore (Billerica, MA) |
|  |  | Strem Chemicals (Newburyport, MA) |
|  |  | ChemGenes (Wilmington, MA) |
|  |  | Santa Cruz Biotechnology (Dallas, TX) |
1.2 Solid-phase RNA synthesis, deprotection, and purification

RNA oligonucleotides P, T₅, and H (See Section S4, page S17) were synthesized by standard solid-phase phosphoramidite polymerization chemistry on a MerMade 6 RNA/DNA oligonucleotide synthesizer. Cleavage and elution of protected oligonucleotide products from 50 µmol universal CPG-solid support columns were performed by equilibrating and eluting the solid support material with a 1:1 mixture of concentrated ammonium hydroxide and concentrated aqueous methylamine for a total of 3 times (equilibration time: 3 x 10 min; elution volume: 3 x 5 mL for 50 µmol columns). Deprotessions of nucleobases and phosphates were carried out by heating the strongly alkaline eluent for 2.5 h at 65 °C; the resultant clear (or off-white) homogeneous mixtures were first evaporated in vacuo for 3 h at 40 °C 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 dimethyl sulfoxide (DMSO) and 2.5 mL of trimethylamine trihydrofluoride (TEA-3HF; CAUTION: TOXIC), and heated for 2.5 h at 65 °C to remove the TBDMS protecting groups of the ribose 2'-hydroxyl groups. The mixtures appeared homogeneous and with pale to golden yellow color. After cooling to room temperature (~30 min), mixtures of 625 µL of 3 M sodium acetate and 15 mL of 1-butanol were used for RNA precipitation. The precipitates were spun down (4000 rpm, 5 min) 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 (Agilent Technologies, Santa Clara, CA), using 25 mM triethylammonium bicarbonate in H₂O (TEAB buffer, pH 7.5) with an increasing gradient of 0 to 15 % acetonitrile over 30 min (flow rate: 10 mL min⁻¹). Elution of RNA was monitored by UV absorption at 254 and 280 nm. The desired RNA fractions were pooled and lyophilized to afford a fluffy white powder. The resultant white residues were again purified by preparative-scale strong anion-exchange HPLC, using the ThermoFisher-Dionex DNAPac PA100 strong anion exchange column (Waltham, MA), with an increasing gradient of 0 to 100 mM aqueous sodium perchlorate solution over 30 min with a flow rate of 10 mL min⁻¹. The desired RNA fractions were collected, pooled and lyophilized 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.
1.3 Characterization of small molecules and oligonucleotides with liquid chromatography high resolution - mass spectrometry (LC-HRMS)

The monoisotopic mass \( (m/z) \) and purity for all synthesized RNA oligomers and small molecules were assessed using an Agilent 1200 high-performance liquid chromatography system coupled to an Agilent 6230 time-of-flight mass spectrometer (Santa Clara, CA). The LC-MS setup was equipped with a solvent degasser, temperature controlled auto sampler, column oven, diode-array detector, and a dual electrospray ionization source. For all analyses, 100–200 pmole of sample was analyzed with extended dynamic range using these settings: column temperature, 50 °C; scan rate, 1 spectrum s\(^{-1}\); mass range, 239 m/z – 3200 m/z; drying gas flow, 8 L min\(^{-1}\); drying gas temperature, 325 °C; nebulizer pressure, 30 psig; capillary voltage, −3500 V; fragmentor, 200 V; and skimmer, 65 V.

1.3.1 Negative ion mode analysis

- Column: Waters 100 mm XBridge C18 column, 1 mm i.d., 3.5 µm particle size (Waters Corporation, Milford, MA).
- Solvent Conditions: Solvent A was water with 200 mM HFIP, and 1.25 mM TEA at pH 7.0. Solvent B was methanol, and was ramped from 2.5% to 20% over 30 min at a flow rate of 0.1 mL/min.\(^1\)

1.3.2 Positive ion mode analysis

- Column: Agilent ZORBAX 300 Rapid Resolution HD SP-C18 column, 50 mm × 2.1 mm, 1.8 µm particle size.
- Solvent Conditions: Solvent A was 0.1% formic acid in water. Solvent B was 0.1% formic acid in acetonitrile, and was ramped from 10% to 100% over 30 min at a flow rate of 0.2 mL/min.

Data analysis was performed using Agilent MassHunter Qualitative Analysis software.

1.4 Determination of mono- and oligonucleotide concentrations

Concentrations of the aqueous mono- and oligonucleotide 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 oligonucleotides at 260 nm were calculated with the OligoAnalyzer software provided by Integrated DNA Technologies (version 3.1, Coralville, IA).\(^2,3\) The extinction coefficients of 2-MelImpddC and 2-MelmpC were assumed to be 7.07 L mmol\(^{-1}\) cm\(^{-1}\), while that of 2-MelmpG and its nonhydrolyzable analogues were assumed to be 12.08 L mmol\(^{-1}\) cm\(^{-1}\).\(^3\)
2. Synthesis of ICG and 2-MelmpddC

2.1 General considerations

Unless otherwise noted, all reactions were conducted in oven-dried round-bottomed flasks that were fitted with appropriately-sized rubber septa, under a positive pressure of dry argon. Aldrich SureSeal™ anhydrous solvents were used for all reactions. Reaction progress was monitored either by means of thin-layer chromatography (TLC) or low-resolution electrospray ionization ion-trap mass spectrometry, on a Bruker Daltonics Esquire 6000 mass spectrometer (Billerica, MA). TLC plates were visualized by 254 nm ultraviolet light, alkaline potassium permanganate stain, or iodine/sand mixture. Normal- and reverse-phase flash chromatography were carried out on a Teledyne Isco CombiFlash Rf system (Lincoln, NE) on pre-packed silica or C18Aq columns.

2.2 NMR data for ICG, 2-MelmpddC and their synthetic intermediates

NMR spectra were recorded on a Varian Inova 400 MHz spectrometer equipped with a broadband PFG (z-gradient) probe (400 MHz for $^1$H, 100 MHz for $^{13}$C, 161 MHz for $^{31}$P; Santa Clara, CA). Proton, carbon and phosphorus chemical shifts were reported in parts per million (ppm) values on the $\delta$ scale, and referenced to residual protium in the NMR solvents (Proton NMR: CHCl$_3$, $\delta$ = 7.26 ppm; DHO, $\delta$ = 4.79 ppm; CHD$_2$OD, $\delta$ = 1.94 ppm; Carbon NMR: CDCl$_3$, $\delta$ = 77.16 ppm; CD$_3$OD, $\delta$ = 49.0 ppm, DMSO-$d_6$: $\delta$ = 39.52 ppm),$^4$ while proton-decoupled phosphorus chemical shifts were internally referenced to trimethyl phosphate (CDCl$_3$: $\delta$ = 3.0 ppm; D$_2$O: $\delta$ = 3.8 ppm; CD$_3$OD: $\delta$ = 3.4 ppm; DMSO-$d_6$: $\delta$ = 3.5 ppm).$^5$ All NMR spectra were recorded at 25 °C. Data were reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, q = quartet, m = multiplet, br = broad), and integration. NMR spectra were processed by MestReNova 11 (Mestrelab Research, Santiago de Compostela, Spain).
2.3 Experimental procedures and compound characterizations

2.3.1. 4-bromo-5-methyl-1-phenylsulfonylimidazole (Compound 2)

5-Bromo-4-methyl-1H-imidazole (compound 1, 1 g, 6.21 mmol) was added to an oven-dried round bottom flask and stir bar, followed by addition of anhydrous dichloromethane (DCM, 25 mL, ~0.25 M). To the partially dissolved starting material in DCM (as a white suspension) was added 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU) (1.04 mL, 6.83 mmol, 1.1 equiv., ρ = 1.018 g mL⁻¹) at ambient temperature, after which the suspension promptly clarified. Afterwards, benzenesulfonyl chloride (1.03 mL, 8.07 mmol, 1.3 equiv., ρ = 1.384 g mL⁻¹) was added dropwise. The reaction was stirred for 30 minutes at ambient temperature, after which the reaction turned bright yellow. The crude reaction mixture was partially evaporated in vacuo, diluted with ethyl acetate (EtOAc), and extracted with water for three times. The organic layer was dried over sodium sulfate and evaporated in vacuo to afford crude solid white residues with a yellow tint. The crude residue was purified by normal-phase flash chromatographic purification on a 40 g pre-packed silica column over 11 column volumes (CV) of 0–30% EtOAc in hexanes. Titled compound 2 was eluted at ~18% EtOAc in hexanes, followed by evaporation in vacuo to afford a white solid residue (900 mg, 2.99 mmol, 48%).

\(^1\)H NMR (CDCl₃) δ 8.05 (s, 1H), 7.94 – 7.87 (m, 2H), 7.77 – 7.68 (m, 1H), 7.65 – 7.56 (m, 2H), 2.25 (s, 3H).

\(^{13}\)C NMR (CDCl₃) δ 137.35, 136.38, 135.29, 130.10, 127.79, 125.34, 118.50, 10.12.

TLC (UV; KMnO₄) Rₚ: 0.30 (20% EtOAc in Hexanes)

HRMS Calc’d for C₁₀H₁₀BrN₂O₂S [M+H]⁺: 300.9646; Found: 300.9648

2.3.2 Diethyl (5-methyl-1-(phenylsulfonyl)-imidazol-4-yl)phosphonate (Compound 3)

To a rigorously flame-dried round bottom flask and stir bar was added compound 2 (530 mg, 1.76 mmol), fresh tetrakis(triphenylphosphine)palladium(0) (305 mg, 0.264 mmol, 0.15 equiv.), and triphenylphosphine (PPh₃, 692 mg, 2.64 mmol, 1.5 equiv.). The flask was then evacuated
and exchanged with dry argon three times. In a separate rigorously flame-dried pear-shaped flask, anhydrous dimethyl sulfoxide (18 mL, ~0.1 M), diethyl phosphate (490 µL, ~3.52 mmol, 2 equiv., ρ = 1.072 g mL⁻¹) and anhydrous triethylamine (TEA, 490 µL, ~3.52 mmol, 2 equiv., ρ = 0.726 g mL⁻¹) were sparged by a stream of dry argon for 30 min. The argon-purged reagent-solvent mixture was then cannula-transferred to the round bottom flask containing the starting material and reagents, followed by 2 h of heating at 110 °C. The reaction mixture slowly homogenized upon heating to form a golden yellow mixture. After 2 h, the reaction was allowed to cool to ambient temperature, and diluted with 150 mL of EtOAc. The organic layer was washed with water and brine, dried over sodium sulfate and evaporated in vacuo. The resulting crude syrup was purified by normal-phase flash chromatography on a 40 g pre-packed silica column over 14 CV of solvents, with a linearly increasing gradient of 0–10% MeOH in CH₂Cl₂. 288 mg of starting material remained unreacted (54.3%), and the titled compound 3 was eluted from the column at ~6% MeOH in DCM, and evaporated in vacuo to afford a colorless syrup (224 mg, 0.625 mmol, 35%).

\[ ^1H \text{NMR (CDCl}_3) \delta \text{8.19 (d, J = 2.4 Hz, 1H), 7.96 – 7.88 (m, 2H), 7.78 – 7.71 (m, 1H),} \]
\[ 7.63 – 7.59 (m, 2H), 4.23 – 4.02 (m, 4H), 2.56 (d, J = 2.3 Hz, 3H), 1.37 – 1.22 (t, J = 7 Hz, 6H). \]

\[ ^{13}C \text{NMR (CDCl}_3) \delta \text{138.24 (d, J = 22 Hz), 137.57 (d, J = 39 Hz), 137.06, 135.38, 130.06,} \]
\[ 129.80 (d, J = 240 Hz), 127.88, 62.69 (d, J = 5.6 Hz), 16.31 (d, J = 7.7 Hz), 10.31. \]

\[ ^{31}P \text{NMR (CDCl}_3) \delta \text{10.07 (s)} \]

TLC (UV; KMnO₄) \( R_f : 0.34 \) (70% EtOAc in Hexanes)

HRMS (m/z) Calc’d for C₁₄H₂₀N₂O₅PS [M+H]+: 359.0831; Found: 359.0841

2.3.3 \( N^2 \)-isobutyryl-2',3'-diacetylguanosine (4-methyl-1H-imidazol-5-yl)phosphonate, triethylammonium salt (Compound 5)

\[
\begin{align*}
\text{N}^2\text{-isobutyryl-2',3'-diacetylguanosine (4-methyl-1H-imidazol-5-yl)phosphonate, triethylammonium salt (Compound 5)}
\end{align*}
\]
Step 1. Prior to initiating this reaction, a round bottom flask containing compound 3 (370 mg, 1.03 mmol) and a stir bar was thoroughly dried under high vacuum overnight. Afterwards, anhydrous DCM (20.6 mL, 0.05 M) and TEA (360 µL, 2.58 mmol, 2.5 equiv., \( \rho = 0.726 \text{ g mL}^{-1} \)) were added to the reaction vessel to dissolve 3 into a colorless homogeneous solution. Bromotrimethylsilane (TMSBr, 680 µL, \( \approx 5.17 \) mmol, 5 equiv., \( \rho = 1.16 \text{ g mL}^{-1} \)) was then added dropwise at ambient temperature, and the reaction mixture was allowed to stir at room temperature overnight; a dense white smoke was observed when bromotrimethylsilane was added. The reaction mixture gradually changed from colorless to a deep and translucent brown color. To quench the excess bromotrimethylsilane and to decompose the resultant bis(trimethylsilyl)phosphonate esters, an excess of methanol was added to the reaction mixture, and the pH of the mixture promptly dropped to \(< 1\) (crudely determined by pH paper). Solvents were evaporated in vacuo, and the residual volatile impurities and hydrogen bromide were azeotropically removed by co-evaporation with toluene for three times. The resultant orange-red solid residue (compound 4) was carried forward to step 2 without further purification.

Step 2. Prior to initiating this reaction, the crude material (compound 4, 100% conversion assumed; 3 equiv.) obtained from step 1 was dried in vacuo under high vacuum overnight. The round bottom flask containing 4 was sequentially charged with \( \text{N}^2\text{-isobutyryl-2',3'-diacetylguanosine} \) (151 mg, 0.344 mmol, 1 equiv.), \( \text{PPh}_3 \) (1.36 g, 5.17 mmol, 15 equiv.) and anhydrous DCM (21 mL, \( \approx 0.05 \) M) under an argon atmosphere. The reaction mixture was cloudy and reddish brown. Using an ice bath, the reaction was allowed to equilibrate to 0 °C for 5 min, after which diisopropyl azodicarboxylate (DIAD; 1.04 mL, 5.17 mmol, 15 equiv., \( \rho = 1.027 \text{ g mL}^{-1} \)) was added dropwise to the reaction mixture over a duration of 10 min. The reaction mixture was observed to clarify into a translucent brown mixture as DIAD was added. The reaction was then allowed to stir for 3 h under ambient temperature. Subsequently, the reaction was stopped by evaporating the solvent in vacuo to form a brown syrup. The residue was crudely purified by normal phase flash chromatography with 12 CV of 0–20% methanol in DCM on a 40 g pre-packed silica column, followed by elution of crude compound 5 in 100% MeOH. The crude material was further purified on a prepacked 50 g C18Aq column, over 20 CV of 0–20% CH₃CN in TEAB. The titled compound 5 was eluted from the column at \( \approx 17.5\% \) CH₃CN in TEAB, followed by lyophilization to afford 5 as a white fluffy solid. Yield was not determined at this step due to the large excess of triethylammonium bicarbonate in the sample. (Note: To afford pure compound 5, the eluted purified product should be promptly lyophilized. The acetyl protecting groups on the 2'- and 3'-hydroxyl groups were found to be labile under basic aqueous conditions.)
\[ ^1H \text{ NMR (DMSO-}\text{d}_6) \delta 8.20 (s, 1H), 7.42 (d, J = 1.8 \text{ Hz}, 1H), 6.08 – 5.95 (m, 2H), 5.46 – 5.37 (m, 1H), 4.25 – 4.22 (m, 1H), 3.98 – 3.87 (m, 1H), 3.80 (dt, J = 10.7, 4.1 \text{ Hz}, 1H), 2.92 (septet, J = 7.0 \text{ Hz}, 1H), 2.72 (\text{TEAH}^+, q, J = 7.1 \text{ Hz}), 2.23 (s, 3H), 2.06 (s, 3H), 1.93 (s, 3H), 1.08 (d, J = 7.0 \text{ Hz}, 3H), 1.05 (d, J = 7.0 \text{ Hz}, 3H), 1.01 (\text{TEAH}^+, t, J = 7.1 \text{ Hz}). \]

\[ ^{13}C \text{ NMR (DMSO-}\text{d}_6) \delta 180.94, 169.46, 169.04, 154.91, 148.63, 148.21, 139.60, 139.14 (\text{br s}), 134.96 (d, J = 15 \text{ Hz}), 122.99 (d, J = 201 \text{ Hz}), 121.05, 85.88, 82.64 (d, J = 8.5 \text{ Hz}), 71.67 (d, J = 14 \text{ Hz}), 62.97 (d, J = 4.1 \text{ Hz}), 45.41 (\text{TEAH}^+), 34.38, 20.49, 20.16, 18.99, 18.83, 12.97, 9.77 (\text{TEAH}^+). \]

\[ ^{31}P \text{ NMR (CD}_3\text{OD)} \delta 5.66 (s) \]

HRMS Calc'd for C\(_{22}\)H\(_{29}\)N\(_7\)O\(_{10}\)P [M+H]\(^+\): 580.1714; Found: 582.1728

### 2.3.4 Guanosine (4-methyl-1H-imidazol-5-yl)phosphonate, sodium salt (ICG)

![Chemical structure of Compound 5 and ICG](image)

Compound 5 was charged to a heavy-walled high-pressure borosilicate glass reaction vessel along with a stir bar, and was then dissolved in a minimum amount of methanol. Afterwards, an excess of concentrated ammonium hydroxide (~20 mL) was added to the vessel. The vessel was sealed, heated to 65 °C with a pre-equilibrated oil bath, and stirred for 4 hours (CAUTION: PRESSURIZED CONTENTS) Afterwards, the reaction was allowed to equilibrate to room temperature (~0.5–1 h), followed by solvent evaporation in vacuo to afford a white solid deposit. The resulting crude material was resuspended in 250 mM TEAB, and the resulting suspension was filtered through a small cotton plug, and purified on a prepacked 50 g C18Aq column, over 20 CV of 0–15% CH\(_3\)CN in TEAB. ICG was eluted from the column at ~5% CH\(_3\)CN in TEAB, followed by lyophilization to afford the titled compound (in triethylammonium cation form) as a white fluffy solid. Sodium cation exchange was performed on the sample, by dissolving the sample in a minimum amount of deionized water, followed by re-precipitation in >25-fold excess (in volume) of 1:1:0.04 (v/v/v) acetone/diethyl ether/saturated NaClO\(_4\) in acetone. The precipitate was centrifuged and washed with acetone twice. Final drying of the centrifuged and washed sample on high-vacuum furnished 34 mg (0.076 mmol) of ICG (Na\(^+\) cation form) as fine
white granular solids. Final percentage yield of ICG was calculated by dividing the number of moles of isolated ICG•Na⁺ by that of \( N^2\)-isobutyl-2′,3′-diacetylguanosine (Section S2.3.3, Pages S7-S9), and was determined to be 22% over three steps.

\[ \text{ICG} \]

\[ \text{N}^2\text{-isobutyl-2',3'-diacetylguanosine (Section S2.3.3, Pages S7-S9)} \]

\[ \text{To an oven-dried round-bottomed flask was added zalcitabine (2',3'-dideoxycytosine, 410 mg, 1.94 mmol) and anhydrous pyridine (pyr., 81 mL, 0.024 M). To the non-homogeneous mixture was then added chlorotrimethylsilane (TMSCl, 1.73 mL, 13.6 mmol, 7 equiv., \( \rho = 0.856 \text{ g mL}^{-1} \)) at room temperature, followed by rigorous stirring and ultrasound sonication for 20 min. The non-homogeneous mixture clarified. Benzoyl chloride (BzCl, 1.24 mL, 10.7 mmol, 5.5 equiv., \( \rho = 1.211 \text{ g mL}^{-1} \)) was then added dropwise over a period of 5 min at room temperature, and the mixture was stirred for 2 h under ambient temperature. The homogeneous reaction mixture turned light yellow. Afterwards, the reaction was cooled to 0 °C in an ice-water bath, followed by dropwise addition of water (12 mL) and concentrated aqueous ammonia (30 mL). The quenched reaction mixture was stirred at room temperature for 30 min, followed by solvent removal via extended rotary evaporation (\( > 30 \text{ min} \)). Half of the final residue was resuspended in a minimum amount of 2:1 water:DMSO, followed by direct reverse-phase flash chromatography on a prepacked 50 g C18Aq column, over 15 CV of 0–40% CH₃CN in TEAB. The titled compound was eluted from the column at \( \sim 30\% \text{ CH}_3\text{CN in TEAB. Solvent evaporation in vacuo of all} \]

\[ \text{2.3.5 N-benzyol-2',3'-dideoxycytosine}^{12} \]

\[ 1. \text{TMSCl, pyr. rt, 30 min.} \]

\[ 2. \text{BzCl, 2h} \]

\[ 3. \text{H}_2\text{O, NH}_4\text{OH 0 °C to rt, 30 min} \]
relevant fractions afforded the titled compound A (400 mg; 1.27 mmol, 65.4 %) as a white fluffy solid.

**H NMR (DMSO-\(d_6\))**
\[\delta 8.55 (d, J = 7.4 \text{ Hz}, 1\text{H}), 8.01 (d, J = 7.7 \text{ Hz}, 2\text{H}), 7.66 - 7.56 (m, 1\text{H}), 7.50 (t, J = 7.6 \text{ Hz}, 2\text{H}), 7.33 (d, J = 7.4 \text{ Hz}, 1\text{H}), 5.97 (dd, J = 6.8, 2.4 \text{ Hz}, 1\text{H}), 4.12 (ddt, J = 9.4, 6.4, 3.5 \text{ Hz}, 1\text{H}), 3.78 (dd, J = 12.1, 3.4 \text{ Hz}, 1\text{H}), 3.60 (dd, J = 12.1, 3.7 \text{ Hz}, 1\text{H}), 2.40 (ddt, J = 14.2, 10.7, 7.4 \text{ Hz}, 1\text{H}), 2.01 (ddt, J = 13.4, 7.3, 2.9 \text{ Hz}, 1\text{H}), 1.93 - 1.70 (m, 2\text{H}).

**\(^{13}\)C NMR (DMSO-\(d_6\))**
\[\delta 167.35, 162.83, 154.42, 145.07, 133.24, 132.66, 128.43, 95.41, 87.02, 82.79, 61.56, 61.47, 32.93, 24.00.

**HRMS**
Calc'd for C\(_{20}\)H\(_{34}\)N\(_8\)O\(_7\)P [M+H]\(^+\): 316.1297; Found: 316.1295

### 2.3.6 \(N\)-benzoyl-2′,3′-dideoxycytosine 5′-phosphoro-2-methylimidazolide, triethylammonium salt\(^{13}\)

![Chemical structure](image)

To a rigorously flame-dried and argon-exchanged round-bottomed flask was added compound A (100 mg, 0.317 mmol) and anhydrous trimethyl phosphate (PO(OMe)_3, 6.4 mL, 0.05 M; Note: PO(OMe)_3 was dried over activated 4 Å molecular sieves overnight). Compound A partially dissolved in PO(OMe)_3 to form a translucent mixture. Diisopropylethylamine (DIPEA, 85 µL, 0.48 mmol, 1.5 equiv., \(\rho = 0.742 \text{ g mL}^{-1}\)) was added, followed by slow, dropwise addition of phosphorus oxychloride (POCl\(_3\), 35 µL, 0.35 mmol, 1.1 equiv., \(\rho = 1.645 \text{ g mL}^{-1}\)) at room temperature. A tinge of golden yellow color could be observed at the end of reagent addition. The solution was stirred at room temperature for 15 minutes, followed by two more rounds of DIPEA and POCl\(_3\) additions. The reaction mixture appeared golden yellow. Afterwards, the reaction mixture was cooled to 0 °C in an ice-water bath, followed by addition of 2-methylimidazole (2-Melm, 130.1 mg, 1.585 mmol, 5 equiv.). The reaction mixture was stirred at room temperature for 15 minutes. The reaction was cooled to 0 °C again before the addition of 500 mM aqueous TEAB (24 mL) as the final quenching reagent, after which the mixture clarified into a lightly yellowish solution. Effervescence was observed upon TEAB addition. The crude mixture was degassed by sonication, and purified by reverse-phase flash chromatography on a
prepacked 50 g C18Aq column, over 25 CV of 0–25% CH₃CN in TEAB. The titled compound was eluted from the column at ~18% CH₃CN in TEAB. Relevant fractions were pooled and lyophilized to afford the titled compound B as a white solid.

**¹H NMR (D₂O)**

δ 8.38 (d, J = 7.5 Hz, 1H), 8.03 – 7.99 (m, 2H), 7.65 – 7.60 (m, 1H), 7.54 – 7.49 (m, 3H), 7.31 (d, J = 7.4 Hz, 1H), 7.14 (t, J = 1.5 Hz, 1H), 6.75 (t, J = 1.5 Hz, 1H), 5.95 (dd, J = 6.7, 2.8 Hz, 1H), 4.15 (dt, J = 9.3, 5.1 Hz, 1H), 3.89 (ddd, J = 11.6, 5.9, 3.2 Hz, 1H), 3.76 (ddd, J = 11.4, 6.8, 4.3 Hz, 1H), 2.98 (TEAH⁺, q, J = 7.3 Hz, 7H), 2.45 (s, 3H), 2.41 – 2.33 (m, 1H), 1.97 (ddt, J = 13.4, 7.7, 3.1 Hz, 1H), 1.87 (ddd, J = 12.4, 7.6, 5.8, 3.2 Hz, 1H), 1.75 – 1.62 (m, 1H), 1.14 (TEAH⁺, t, J = 7.3 Hz, 11H)

**¹³C NMR (D₂O)**

δ 167.40, 162.85, 154.32, 146.39 (d, J = 4.2 Hz), 144.93, 133.27, 132.64, 128.42, 124.36 (d, J = 9.5 Hz), 121.92 (d, J = 5.2 Hz), 120.94, 95.63, 87.03, 80.67 (d, J = 8.0 Hz), 65.35 (d, J = 6.1 Hz), 45.52 (TEAH⁺), 32.62, 24.31, 14.57, 8.93 (TEAH⁺).

**³¹P NMR (D₂O)**

δ –8.85 (s)

**HRMS**

Calc’d for C₂₀H₃₄N₈O₇P [M–H]⁻: 458.1235; Found: 458.1245

---

2.3.7 2',3'-dideoxycytosine 5'-phosphoro-2-methylimidazolide, triethylammonium salt (2-MelmpddC)

**[Diagram]**

Compound B was charged to a heavy-walled high-pressure borosilicate glass reaction vessel equipped with a stir bar, and was dissolved in a minimum amount of methanol. Afterwards, an excess of concentrated ammonium hydroxide (28% NH₃ in H₂O; ~10 mL) was added to the vessel. The vessel was sealed, heated to 65 °C with a pre-equilibrated oil bath, and stirred for 3 hours (CAUTION: PRESSURIZED CONTENTS) Afterwards, the reaction was allowed to equilibrate to room temperature (~0.5–1 h), followed by solvent evaporation in vacuo to afford a white solid deposit. The resulting crude material was resuspended with 250 mM TEAB and purified on a prepacked 50 g C18Aq column, over 20 CV of 0–15% CH₃CN in TEAB. The titled
compound was eluted from the column at ~5% CH$_3$CN in TEAB. Relevant fractions were pooled and lyophilized to afford the titled compound (in triethylammonium cation form) as a white solid.

|   |   |
|---|---|
| **$^1$H NMR (D$_2$O)** | δ 7.73 (dd, $J = 7.4$, 1.3 Hz, 1H), 7.08 (q, $J = 1.3$ Hz, 1H), 6.66 (t, $J = 1.5$ Hz, 1H), 5.92 (dd, $J = 6.6$, 3.8 Hz, 1H), 5.70 (d, $J = 7.4$ Hz, 1H), 4.03 (dq, $J = 7.5$, 3.9 Hz, 1H), 3.78 (ddd, $J = 11.5$, 5.8, 3.3 Hz, 1H), 3.68 (ddd, $J = 11.3$, 6.6, 4.6 Hz, 1H), 2.84 (TEAH$^+$, q, $J = 7.2$ Hz, 8H), 2.40 (d, $J = 1.2$ Hz, 2H), 2.24 – 2.15 (m, 1H), 1.82 (m, 2H), 1.66 (dt, $J = 12.3$, 8.3 Hz, 1H), 1.08 (TEAH$^+$, t, $J = 7.2$ Hz, 11H). |
| **$^{13}$C NMR (D$_2$O)** | δ 165.64, 155.28, 146.48 (d, $J = 4.2$ Hz), 140.74, 125.29 (d, $J = 10$ Hz), 121.75 (d, $J = 6.1$ Hz), 93.59, 85.80, 79.29 (d, $J = 8.4$ Hz), 65.89 (d, $J = 6.0$ Hz), 45.56 (TEAH$^+$), 32.11, 25.05, 14.86, 9.53 (TEAH$^+$) |
| **$^{31}$P NMR (D$_2$O)** | δ –8.72 (s) |
| **HRMS** | Calc’d for C$_{20}$H$_{34}$N$_8$O$_7$P [M–H]$: 354.0973$; Found: 354.1001 |

3. $pK_a$ determination of 2-MelmpG and nonhydrolyzable analogues

3.1 Basic principles

We measured the $pK_a$ of 2-MelmpG or its nonhydrolyzable analogue by acid titration. Protonation of the basic nitrogen atoms on 2-methylimidazole or the heteroaryl-phosphonate mimics causes the ring to be formally positively charged, shifting proton resonances downfield. The chemical shifts of these aromatic protons were plotted against the corresponding pH values to afford a sigmoidal curve. At the inflection point in the Henderson-Hasselbach equation,

$$\text{pH} = \text{p}K_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

the ratio between the monomer and its protonated conjugate acid is 1:1, yielding the $pK_a$ value.

In our NMR experiments, monomer preparations and pH adjustments were performed using D$_2$O and deuterated reagents, both for NMR locking and to minimize the HOD peak (4.79 ppm) on the proton spectrum. To correct the obtained $pK_a$ values to their H$_2$O equivalent, we calculate:

$$pK_a^{\text{H}_2\text{O}} = 0.929 \cdot pK_a^{\text{D}_2\text{O}} + 0.42$$
3.2 Experimental conditions

30–40 mM solutions of 2-MelmpG and its analogues in D₂O, 5 M sodium chloride in D₂O and 1 M magnesium chloride in D₂O solutions were prepared. The final sample solutions for NMR spectral acquisition, with a sample scale of 400 µL, contained 15 mM monoribonucleotide, 500 mM NaCl, and 100 mM MgCl₂. The pD of the final analyte solutions were adjusted by either sodium deuteroxide or deuterium chloride in D₂O. All proton spectra were acquired with 256 scans, 1 s relaxation delay, and 45° pulse angle at 25 °C. Proton chemical shifts were referenced to residual protium in D₂O (δ = 4.79 ppm).

For 2-MelmpG, ICG, and PYG, the pD of these solutions were first adjusted to ~10; for PZG, pD was adjusted to ~5. The analyte solutions were gradually acidified using DCI/D₂O solutions, followed by spectral acquisitions. The chemical shift of the aromatic protons highlighted in red (Figures S1–3, page S15–S16) was traced. For 2-MelmpG, ICG and PZG, sample acidification and spectral acquisition were repeated until the samples were at a pD value at which the reporter proton did not undergo further downfield shift. For PYG, acidification and spectral recording were repeated until the sample was at pD = 0.7.

3.3 Derivation of the pKᵦ fitting equation

We assume that proton exchange on the leaving group moieties of the monomers or analogues is faster than the NMR time scale. Under this scenario, the observed chemical shift of the proton of interest is a weighted average between that of the unprotonated monomer (A⁻) and its protonated conjugate acid (HA):

\[ \delta_{\text{obs}} = \delta_{HA} X_{HA} + \delta_{A^-} X_{A^-} \]

Wherein mole fractions (X) of HA and A⁻ are given by:

\[ X_{HA} = \frac{[HA]}{[HA] + [A^-]} ; \quad X_{A^-} = \frac{[A^-]}{[HA] + [A^-]} \]

Thus:

\[ \delta_{\text{obs}} = \frac{\delta_{HA}[HA] + \delta_{A^-}[A^-]}{[HA] + [A^-]} \quad (1) \]

The concentration of the unprotonated monomer, [A⁻], at any pH is dictated by the Henderson-Hasselbach equation:

\[ \text{pH} = pK_a + \log_{10} \left( \frac{[A^-]}{[HA]} \right) \]

\[ \Rightarrow \quad \text{pH} - pK_a = \log_{10} \left( \frac{[A^-]}{[HA]} \right) \]
\[ \Rightarrow \left[ A^- \right] = 10^{(pH-pK_a)} \]
\[ \Rightarrow [A^-] = [HA] \cdot 10^{(pH-pK_a)} \quad (2) \]

By substituting equation 2 back to equation 1, we get:

\[ \delta_{OBS} = \frac{\delta_{HA} \cdot [HA] + \delta_{A^-} \cdot [HA] \cdot 10^{pH-pK_a}}{[HA] + [HA] \cdot 10^{pH-pK_a}} \]

\[ \delta_{OBS} = \frac{\delta_{HA} + (\delta_{A^-}) \cdot 10^{(pH-pK_a)}}{1 + 10^{pH-pK_a}} \]

### 3.4 Representative data

**Figure S1.** Variable pD-¹H NMR spectroscopy of 2-MelmpG. The heteroaromatic proton of 2-MelmpG labeled with red square (top right), with a more upfield proton chemical shift as predicted by Chemdraw 16 (PerkinElmer Informatics, Waltham, MA), was monitored as the monomer solution was gradually acidified. Experiments were performed in duplicate. **Left:** Representative stacked proton spectra of the aromatic region. The HOD peaks, referenced at 4.79 ppm, are not shown to scale. **Right:** The chemical shifts of the labeled proton were plotted against their respective monomer pD values, followed by fitting to estimate the pKₐ value (6.7 ± 0.1) of the leaving group.
Figure S2. Variable pD-\textsuperscript{1}H NMR spectroscopy of PZG. Experiments were performed in duplicate. **Left:** Representative stacked proton spectra of the aromatic region. The HOD peaks are not shown to scale. **Right:** The chemical shift vs. monomer pD plot. The pK\textsubscript{a} value of the leaving group analogue is estimated to be 1.9 ± 0.1.

Figure S3. Variable pD-\textsuperscript{1}H NMR spectroscopy of PYG. **Left:** Representative stacked proton spectra of the aromatic region. As the monomer solution was acidified beyond ~ pD 2.0, the proton α to the pyrrole NH (labelled by red square, top right) preferentially underwent proton-deuteron exchange prior to peak shifting,\textsuperscript{17} leading to the disappearance of the peak. The proton β to the pyrrole NH (labelled by blue circle) underwent proton-deuteron exchange after the α proton had been significantly exchanged. The HOD peaks are not shown to scale. **Right:** The chemical shift vs. monomer pD plot.
4. Isothermal Titration Calorimetry (ITC)

4.1 Experimental conditions

ITC^{18–20} was used to determine the thermodynamic parameters of binding between 2-MelmpG and its nonhydrolyzable analogues with the P/Ts/H sandwich duplex (Sequences: P (Primer): 5′-CUCAAAUG-3′; Ts (Template): 5′-GAGUUACCAUUGAG-3′; H (Helper), 5′-GUAACUC-3′). RNA strands P, Ts and H were synthesized on a MerMade Synthesizer on a 50 µmol scale in accordance to procedures laid out in Section S1.2. All the ITC analyses were performed on a Nano-ITC calorimeter (TA instruments, New Castle, DE). 170 µL of the cell solution, which contained 1.5 mM each of the P, Ts and H strands, 450 mM NaCl, 50 mM Na•HEPES at pH 7, was loaded into the sample cell of the instrument. Then, 90 µL of the syringe solution, which contained the same concentration of the RNA strands, NaCl and Na•HEPES as the cell solution in addition to 15 mM guanosine nucleotides, was prepared separately and loaded into the injection syringe. The instrument was equilibrated at 20 °C for 1 hour; afterwards, 32 injections of 1.5 µL syringe solution were incrementally added to the sample cell at 300-second intervals. A plot of the resulting raw heat rate (in µcal s⁻¹) versus time (s) was recorded; the plot was then integrated to obtain the heat evolved per injection of syringe solution (in µcal). From this data, the heat evolved per change in monomer concentration (in kcal mol⁻¹) was measured. This was then plotted against the cumulative concentration of monomer throughout the titration, resulting in a sigmoidal curve, which was then fitted against the binding isotherm equation derived in Section S4.3.
4.2 LC-HRMS analysis of oligonucleotides

Figure S4. UV absorbance at 260 nm, and monoisotopic monoanionic mass for the synthesized RNAs for the ITC studies. Based on the UV absorbance traces, the purity was determined to be >90% for the synthesized RNAs. Found: monoisotopic monoanionic mass; Calculated: calculated exact monoanionic mass. Y-axis: milli-absorbance response units at 260 nm.

4.3 Derivation of the ITC binding isotherm

In our proposed mechanism, a molecule of 2-MelmpG or its analogues (M) binds to the P/Ts/H sandwich duplex (D), with a 1:1 ratio, to form DM, and the binding affinity is reflected by equilibrium association constant $K_a$. Throughout the ITC experiments, monomers or analogues were incrementally added to the P/Ts/H duplex. The total heat change resultant from RNA-ligand binding ($Q$) is a product of the molar enthalpy of duplex-monomer binding ($\Delta H$), the volume of the ITC cell ($V_0$), and the concentrations of monomer-bound complex DM ([DM]). Thus:

$$Q = \Delta H \cdot V_0 \cdot [DM] \quad (1)$$

Since both $\Delta H$ and $V_0$ are constants, change in heat throughout the titration ($dQ$) must be the result of the change in concentration of DM ($d[DM]$):

$$dQ = \Delta H \cdot V_0 \cdot d[DM] \quad (1a)$$
Ultimately, both $dQ$ and $d[DM]$ should be related to the change in total concentration of the monomer ($d[M_{tot}]$). We can therefore differentiate both sides of the equation with respect to $[M_{tot}]$, and we get:

$$\frac{dQ}{dM_{tot}} = \Delta H \cdot V_o \cdot \frac{d[DM]}{dM_{tot}} \quad (2)$$

To express $[DM]$ in terms of $[M_{tot}]$, we need to consider the following equilibrium, which holds if and only if the binding stoichiometry between the RNA duplex and the monomer is 1:1,

$$K = \frac{[DM]}{[D_{free}] \cdot [M_{free}]} \quad (3)$$

Upon addition of monomer, some portion of total duplex ([D$_{tot}$]) will associate with a portion of added monomer to form [DM], while some will remain in the unbound state ([D$_{free}$]). The same applies to the monomer. The total concentration of duplex and monomer at any point throughout the titration is given by:

$$[D_{free}] + [DM] = [D_{tot}] \quad ; \quad [M_{free}] + [DM] = [M_{tot}] \quad (4)$$

Equations (3) and (4), when combined, give:

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

$$[DM]^2 - \left( M_{tot} + D_{tot} + \frac{1}{K} \right) [DM] + M_{tot}D_{tot} = 0 \quad (5)$$

Hence, the two roots (solutions) of equation 5 are:

$$[DM] = \frac{\left( M_{tot} + D_{tot} + \frac{1}{K} \right) \pm \left[ \left( M_{tot} + D_{tot} + \frac{1}{K} \right)^2 - 4M_{tot}D_{tot} \right]^{\frac{1}{2}}}{2} \quad (6)$$

One of the two roots must be rejected to arrive at the proper isotherm. When $K$ approaches positive infinity (\( \lim_{K \to \infty} [DM] \)), 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:

$$[DM] = \frac{\left( M_{tot} + D_{tot} + \frac{1}{K} \right) - \left[ \left( M_{tot} + D_{tot} + \frac{1}{K} \right)^2 - 4M_{tot}D_{tot} \right]^{\frac{1}{2}}}{2} \quad (6a)$$

From this equation, it is apparent that the change in concentration of the duplex-monomer complex ($d[DM]$) is indeed related to the change in total monomer concentration ($dM_{tot}$).
Subsequently:

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

By incorporating equation 7 into equation 2, we get the ITC binding isotherm (equation 8).

\[
\frac{dQ}{dM_{tot}} = \frac{\Delta H \cdot V_0}{2} \cdot \left\{ 1 - \frac{\left( M_{tot} - D_{tot} + \frac{1}{K} \right)}{\left[ \left( M_{tot} + D_{tot} + \frac{1}{K} \right)^2 - 4M_{tot}D_{tot} \right]^{\frac{1}{2}}} \right\}
\]

At this stage, the stoichiometric factor \( n \) is introduced to the term \( D_{tot} \) to reflect the binding stoichiometry between the RNA sandwich duplex and the ligand; in the ideal scenario, \( n = 1 \). To account for any heat that originates from non-idealities, e.g. heats of dilution, a constant \( a \) is added, yielding the final form:

\[
\frac{dQ}{dM_{tot}} = \frac{\Delta H \cdot V_0}{2} \cdot \left\{ 1 - \frac{\left( M_{tot} - nD_{tot} + \frac{1}{K} \right)}{\left[ \left( M_{tot} + nD_{tot} + \frac{1}{K} \right)^2 - 4nM_{tot}D_{tot} \right]^{\frac{1}{2}}} \right\} + a
\]
4.4 Representative data

Figure S5. (a–e) Plots of the change of heat evolved per change in monomer/analogue concentration of one injection ($\Delta q \cdot \Delta[\text{analogue}]^{-1}$), versus the cumulative change in monomer/analogue concentration, for (a) 2-MelmpG, (b) ICG, (c) ICG with 50 mM Mg$^{2+}$, (d) PZG and (e) PYG. During an ITC titration run, 50 µL of syringe solution, which contained 15 mM monomer/analogue, 1.5 mM of P/Th, 50 mM pH 7 Na•HEPES and 450 mM NaCl, was titrated into 170 µL of cell solution, which did not contain the analogue. All titrations were run in triplicate.
| Ligand         | Fitted Parameters | Derived Parameters |
|---------------|------------------|-------------------|
|               |                  | $K_a \quad (M^{-1})$ | $\Delta H \quad (kcal \: mol^{-1})$ | $n$ | $\Delta G \quad (kcal \: mol^{-1})$ | $\Delta S \quad (cal \: mol^{-1} \: K^{-1})$ |
|               |                  |                   |                 |   |                   |                  |
| 2-MeImpG      | 1                | $2.3 \times 10^3$  | $-16.7$          | 0.75 | $-4.52$         | $-41.5$          |
|               | 2                | $2.4 \times 10^3$  | $-14.3$          | 0.74 | $-4.53$         | $-33.4$          |
|               | 3                | $1.6 \times 10^3$  | $-20.3$          | 0.66 | $-4.31$         | $-54.5$          |
|               | Avg              | $(2.1 \pm 0.4) \times 10^3$ | $-17.1 \pm 3.0$ | 0.72±0.05 | $-4.45 \pm 0.12$ | $-43.1 \pm 10.7$ |
| ICG           | 1                | $3.6 \times 10^3$  | $-14.0$          | 0.81 | $-4.77$         | $-31.5$          |
|               | 2                | $4.0 \times 10^3$  | $-14.2$          | 0.84 | $-4.83$         | $-32.0$          |
|               | 3                | $4.7 \times 10^3$  | $-14.0$          | 0.93 | $-4.92$         | $-30.9$          |
|               | Avg              | $(4.1 \pm 0.5) \times 10^3$ | $-14.1 \pm 0.1$ | 0.86±0.06 | $-4.84 \pm 0.07$ | $-31.5 \pm 0.5$ |
| ICG (w/ 50 mM Mg$^{2+}$) | 1 | $5.6 \times 10^3$ | $-14.8$ | 0.83 | $-5.03$ | $-33.3$ |
|               | 2                | $4.0 \times 10^3$  | $-15.7$          | 0.80 | $-4.83$         | $-37.0$          |
|               | 3                | $3.7 \times 10^3$  | $-15.1$          | 0.72 | $-4.79$         | $-35.1$          |
|               | Avg              | $(4.5 \pm 1.0) \times 10^3$ | $-15.2 \pm 0.5$ | 0.78±0.06 | $-4.89 \pm 0.13$ | $-35.1 \pm 1.9$ |
| PZG           | 1                | $3.1 \times 10^3$  | $-14.8$          | 0.80 | $-4.69$         | $-34.6$          |
|               | 2                | $2.9 \times 10^3$  | $-13.7$          | 0.91 | $-4.64$         | $-31.0$          |
|               | 3                | $2.5 \times 10^3$  | $-14.1$          | 0.93 | $-4.56$         | $-32.4$          |
|               | Avg              | $(2.8 \pm 0.3) \times 10^3$ | $-14.2 \pm 0.6$ | 0.88±0.07 | $-4.63 \pm 0.06$ | $-32.7 \pm 1.8$ |
| PYG           | 1                | $3.8 \times 10^3$  | $-15.1$          | 0.87 | $-4.80$         | $-35.2$          |
|               | 2                | $3.4 \times 10^3$  | $-14.5$          | 0.90 | $-4.74$         | $-33.3$          |
|               | 3                | $3.8 \times 10^3$  | $-14.8$          | 0.86 | $-4.80$         | $-34.3$          |
|               | Avg              | $(3.6 \pm 0.2) \times 10^3$ | $-14.8 \pm 0.3$ | 0.88±0.02 | $-4.78 \pm 0.03$ | $-34.2 \pm 0.9$ |

Table S1. By fitting the ITC data to the binding isotherm derived in Section S5.2 (pages S16-S18), we can obtain the affinity constant ($K_a$), change in enthalpy ($\Delta H$) and stoichiometry factor ($n$) of RNA-analogue binding. Using the relationships $\Delta G = -RT\ln(K_a) = \Delta H - T\Delta S$, we can further obtain the free energy ($\Delta G$) and entropy change ($\Delta S$) of RNA-analogue binding. The adjusted $R^2$ of all fits were > 0.99.
5. Nonenzymatic Primer Extension

5.1 Preparation of cytosine 5'-phosphoro-2-methylimidazolide sodium salt (2-MelmpC•Na⁺)

The protocol employed for the synthesis of 2-MelmpC was adapted from that of Joyce, Inoue and Orgel and our previous report.

Cytosine 5'-monophosphate (CMP) free acid (1 mmole, 1 equiv., 323 mg), 2-methylimidazole (5 mmoles, 5 equiv., 411 mg), triphenylphosphine (5 mmoles, 5 equiv., 1.31 g), anhydrous triethylamine (ρ = 0.726 g mL⁻¹, 5 mmoles, 5 equiv., 700 µL), and 20 mL anhydrous DMSO (50 mM) was sequentially charged into a flame-dried round-bottom flask. The content was heated and sonicated until complete homogenization (~ 10 min), followed by addition of 2,2′-dipyridyl disulfide (5 mmoles, 5 equiv., 1.1 g). The reaction was allowed to stir for 1 h. The crude product was obtained, as an off-white precipitate, by adding the reaction mixture slowly into an ice-cold solution of 200 mL acetone, 200 mL diethyl ether, 18 mL of triethylamine, and 2.5 mL of saturated sodium perchlorate (NaClO₄) in acetone. The precipitate was then collected by suction filtration, followed by volatile removal under high vacuum. The resulting crude residue was purified on a 50 g C18Aq column over 20 CV of of 0–10% CH₃CN in TEAB (pH 7.5). The titled compound was eluted from the column at ~5% CH₃CN in TEAB, followed by lyophilization to afford the titled compound (in triethylammonium cation form) as a white fluffy solid. Lyophilized sample was resuspended in a minimum amount of deionized water, followed by precipitation in >25-fold excess (in volume) of 1:1:0.04 (v/v/v) acetone/diethyl ether/saturated NaClO₄ in acetone. The precipitate was centrifuged and washed with 1:1 acetone/ether twice. Final drying of the centrifuged and washed sample on high-vacuum furnished 2-MelmpC•Na⁺ as a fine white granular solid.
5.2 Primer Extension: Experimental Conditions

Primer extension reactions were set up, in 10 µL scales, with the following components: 1.2 µM Cyanine 3-labeled primer (Sequence: 5’-Cy3-GACUGACUGG-3’), 2.2 µM complementary template (Sequence: 5’-AACCGCCAGUCAGUC-3’), 200 mM Na•HEPES (pH 7 and 8) or Na•MES buffer (pH 5.5 and 6.3), 100 mM MgCl₂, and monomers (Set 1: 30 mM 2-MelmpC; Set 2: 30 mM 2-MelmpC & 30 mM 2MelmpG; Set 3: 30 mM 2MelmpC & 30 mM ICG•Na⁺; Set 4: 30 mM 2-MelmpC & 30 mM PZG•Na⁺; Set 5: 30 mM 2MelmpC & 30 mM PYG•Na⁺; Set 6: 30 mM 2MelmpC & 30 mM GMP•2Na⁺).

Figure S6. Schematic diagram of the primer extension reaction. The 5’-Cy3-labeled primer was annealed with the template in the presence of 100 mM MgCl₂ and 200 mM buffer of desired pH at room temperature. 2-MelmpC (cyan) and stacking 2-MelmpG or nonhydrolyzable analogues (brown) was then added to the primer-template complex. The monomers first bind noncovalently to the primer-template complexes, followed by an irreversible extension step between primer and 2-MelmpC to afford the extended product.

The pH of the stock solutions of both monomers and analogues were adjusted to the pH value at which the primer extensions were run. Primer, template, aqueous MgCl₂ and buffer of the appropriate pH were first combined and thoroughly mixed in 200 µL thin-walled PCR tubes, followed by addition of the appropriate amount of stock solutions of 2-MelmpC and guanosine monomers or analogues (~100 mM, concentration determined by Nanodrop) to initiate the extension reactions. At appropriate time points, 0.5 µL aliquots of these reactions were quenched by addition to 15 µL of a quench buffer containing 1X TBE (89 mM Tris-borate and 2 mM EDTA, overall pH 8.3), 8.0 M urea, and 100 mM pH 8.0 EDTA•4Na⁺. 3 µL of these quenched samples, which contained 0.12 pmole of the fluorescently-labelled primer as well as their extended products, were loaded onto a 200 mm × 200 mm × 0.4 mm denaturing polyacrylamide gel,²⁴ which contained 20% acrylamide monomers with a ratio of 19:1.
acrylamide: $N,N'$-methylenebis(acrylamide), as well as 7.5 M urea (National Diagnostics, Atlanta, GA), followed by electrophoresis at 25 W for 1.5 – 2 h. Reaction rates were calculated by quantifying the conversion of primer to its extended products, and the log values of the fraction of unreacted primer was plotted against time. A linear regression was performed and the negative slope of the fit as plotted was reported as the initial pseudo-first order rate $k_{obs}$.

Imaging and quantification of the fluorescently labelled primer and extended products were performed with an Amersham Biosciences Typhoon 9410 Imager, and the ImageQuant TL software package (GE Healthcare, Little Chalfont, United Kingdom). The $k_{max}$ and $K_M$ of saturation curves (Figure 4C, main text; Figures S13–S14, Pages S31–S32) was estimated by GraphPad Prism 5 (GraphPad Software, San Diego CA).

### 5.3 Representative data

**Figure S7.** Primer extension reactions at pH 5.5. **Top:** Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1: Extended primer+C product. **Bottom:** Plots of the natural log values of the fraction of unreacted primer ($\ln (P/P_{0})$) against time. The linear fits of these plots were used to determine the pseudo first-order observed rate constants ($k_{obs}$). All primer extension experiments presented herein were performed in triplicate, and the standard deviation was presented as error. N/A: rate too low to be reliably fitted.
Figure S8. Primer extension reactions at pH 6.3. Top: Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1, +2: Extended primer+C / primer+CG products, respectively. Bottom: Plots of the natural log values of the fraction of unreacted primer (ln (P/P₀)) against time. All experiments were performed in triplicate.
**Figure S9.** Primer extension reactions at pH 7.0. **Top:** Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1, +2, +3, +4: Extended primer+C / primer+CG / primer+CGG / primer+CGGG products, respectively. **Bottom:** Plots of the natural log values of the fraction of unreacted primer (ln (P/P₀)) against time. All experiments were performed in triplicate.
Figure S10. Primer extension reactions at pH 8.0. **Top:** Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1, +2, +3, +4: Extended primer+C / primer+CG / primer+CGG / primer+CGGG products, respectively. **Bottom:** Plots of the natural log values of the fraction of unreacted primer (ln (P/P₀)) against time. All experiments were performed in triplicate. We also performed primer extension reactions with only 2-MelmpG, in the absence of 2-MelmpC, at pH 5.5, 6.3, 7.0 and 8.0. No extended product was observed after 20 hours (data not shown).
Figure S11. ICG-stacked primer extension reactions at pH 6.3. **Top:** Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1: Extended primer+C product. **Bottom:** Plots of the natural log values of the fraction of unreacted primer (ln (P/P₀)) against time. All experiments were performed in triplicate. The Michaelis-Menten plot of ICG-stacked extensions at pH 6.3 was shown in Figure 4C (main text).
Figure S12. Primer extension reactions in the presence of ICG at pH 8.0. (a) Top: Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1: Extended primer+C product. Bottom: Plots of the natural log values of the fraction of unreacted primer (ln (P/P₀)) against time. All experiments were performed in triplicate. (b) Saturation curve of the ICG-stacked primer extensions at pH 8.0. At this pH, increasing concentrations of ICG do not enhance the rate of primer extension. At rates of ~0.07 h⁻¹, the rates of ICG-stacked extensions are comparable with that of the unstacked (0 mM ICG) extension. Hence, when ICG is used as a stacking monomer at pH 8.0, it offers minimal catalysis to primer extension. Significant precipitation was observed for the primer extension with 60 mM ICG, possibly explaining the slightly lower extension rate.
Figure S13. 2-MelmpG-catalyzed primer extension reactions at pH 6.3. (a) **Top:** Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1, +2, +3: Extended primer+C / primer+CG / primer+CGG products, respectively. **Bottom:** Plots of the natural log values of the fraction of unreacted primer (ln (P/P₀)) against time. All experiments were performed in triplicate. (b) Saturation curve of the 2-MelmpG-stacked primer extensions at pH 6.3. The estimated kₘₐₓ and Kₘ was 0.155 ± 0.008 h⁻¹ and 16 ± 2 mM, respectively (R² = 0.983). At this pH, the rate enhancement afforded by 2-MelmpG approaches the maximum when the concentration of 2-MelmpG is at, or above 30 mM.
Figure S14. 2-MeImpG-catalyzed primer extension reactions at pH 8.0. (a) Top: Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1, +2, +3, +4: Extended primer+C / primer+CG / primer+CGG / primer+CGGG products, respectively. Bottom: Plots of the natural log values of the fraction of unreacted primer (ln \( P/P_0 \)) against time. All experiments were performed in triplicate. (b) Saturation curve of the 2-MeImpG-stacked primer extensions at pH 8.0. The estimated \( k_{\text{max}} \) and \( K_M \) was 17 ± 2 h\(^{-1}\) and 80 ± 18 mM, respectively (\( R^2 = 0.988 \)). Therefore, the 60-fold rate increase conferred by 30 mM 2-MeImpG at primer extension pH of 8.0, as shown in Figure 4B (main text), does not represent the maximal catalytic capability conferred by 2-MeImpG; the true rate enhancement afforded by downstream stacking 2-MeImpG would be greater than 60-fold at pH 8.0. The much higher observed level of extension rate catalysis displayed by 2-MeImpG at pH 8.0 is possibly related to the enhanced formation of the imidazolium-bridged dinucleotide intermediate.
6. CplCG: Detection and Structure Elucidation

6.1 Coincubation of 2-MelmpddC with nonhydrolyzable analogues

6.1.1 Experimental conditions: liquid chromatography, mass spectrometric and tandem MS/MS analyses

2-MelmpddC-analogue coincubation experiments were set up, in 20 μL scales, with the following components: 30 mM nonhydrolyzable analogues, 30 mM 2-MelmpddC, and 200 mM buffer (pH 6.3 Na•MES or pH 8.0 Na•HEPES), followed by prompt pH adjustments with aqueous hydrochloric acid or sodium hydroxide. At appropriate time points, 1 μL aliquots of these reactions were diluted 100-fold with 99 μL of water, followed by prompt injection of 1–2 μL of the diluted samples (300–600 pmole of 2-MelmpddC and analogue) into the liquid-chromatography-coupled mass spectrometer.

The formation of dideoxycytidine-analogue conjugate was assessed using an Agilent 1100 high-performance liquid chromatography system coupled with a ThermoFisher LTQ XL linear ion trap mass spectrometer. The LC-MS setup was equipped with a solvent degasser, autosampler, diode array detector, and an electrospray ionization source. Chromatographic resolution of 2-MelmpddC, analogue, as well as the ddC-G dinucleotide conjugates and degradation products was achieved under reverse-phase conditions across an Agilent Polaris Amide-C18 column (150 mm × 2.0 mm i.d., 3 μm particle size), with aqueous mobile phase being 5 mM neutral triethylammonium acetate buffer and organic mobile phase being acetonitrile. Organic mobile phase was ramped linearly from 2% to 3.2% over 0–6 min and from 3.2% to 7.0% over 6–20 min. The flow rate was maintained at 0.2 mL min⁻¹ and the column was kept under ambient temperature. Concomitantly, the m/z ratio of the chromatographed eluent was continuously monitored in negative ion mode from 150–1000 m/z with a scan rate of 16,667 Da s⁻¹, with the following parameters: drying gas flow, 25 L/min; capillary temperature, 350 °C; spray voltage, 4000 V; spray current, 100 μA; capillary voltage, −0.5 V; and tube lens voltage, −52 V. For the tandem MS/MS analysis, the source parameters were identical, and the parent ions were fragmented using collision induced dissociation (CID) with nitrogen, and normalized collision energy was set at 25.

See Figures S15–17, 19 and 20 (Pages S35–37, S39–40) for representative data.
6.1.2 Experimental conditions: $^{31}$Phosphorus NMR Time Courses

2-MelmpddC and ICG were coincubated in 100 µL scales, with the following components: 30 mM ICG, 30 mM 2-MelmpddC, and 200 mM pH 6.3 Na•MES, and 20% D$_2$O. This sample solution was promptly adjusted to pH 6.3 with minimal amounts of aqueous hydrochloric acid or sodium hydroxide, transferred to a 3mm NMR tube (Wilmad-LabGlass, SP Scienceware, Warminster, PA), followed by $^{31}$P NMR time course analysis over 12 h on a Varian Inova 400 MHz spectrometer equipped with a broadband PFG (z-gradient) probe (161 MHz for $^{31}$P). Proton-decoupled phosphorus chemical shifts were reported in parts per million (ppm) values on the δ scale; each $^{31}$P spectrum was acquired with 512 scans, 1 s relaxation delay, spectral window of –26 to 32 ppm, and 45° pulse angle at 25 °C (Acquisition time per spectrum was approximately 30 min), and acquired at the beginning of each full hour of incubation. At the end of the time course analysis, deuterated Na$_2$HPO$_4$ was introduced to the 2-MelmpddC-ICG sample as an internal reference (final Na$_2$HPO$_4$ concentration ~30 mM). The sample was adjusted to pH 6.3 with minimal aq. HCl and NaOH, and the $^{31}$P spectrum was reacquired to afford a Na$_2$HPO$_4$-referenced spectrum for the last time point. The $^{31}$P resonance of the added Na$_2$HPO$_4$ at pH 6.3 was determined to be 0.87 ppm, using an external reference of 85% H$_3$PO$_4$ (0 ppm by definition) in D$_2$O contained in a coaxial NMR tube. All other $^{31}$P spectra (spectra corresponding to time points 0−11h) were referenced to this Na$_2$HPO4-referenced spectrum.

See Figure S18 (Page S38) for representative data.

*Notes on sample transfer and NMR tube cleaning:* Samples were transferred into the 3 mm NMR tube by a Hamilton 500 µL gas-tight glass syringe and 7-inch, 22s-gauged needle assembly (needle inner diameter: 0.152 mm; Reno, NV); dead volume within the needle was estimated to be 3.2 µL. The syringe-needle assembly was thoroughly washed by Decon Contrad 70 detergent (King of Prussia, PA), 18 Ω water and deuterium oxide prior to use. 3 mm NMR tubes were cleaned by soaking in conc. HCl for >1 h, followed by thorough 18 Ω water and acetone rinsing. Residual acetone was removed by vacuum evaporation inside a house-vacuumed chamber overnight. 3mm NMR tubes should not be dried inside an oven.
6.1.3 Representative data

Figure S15. 2-MelmpddC was coincubated with ICG at pH 6.3. The formation of ddC-ICG dinucleotide was assessed over 13 h. The red, purple, brown, orange and green peaks correspond to ddC-ICG dinucleotide, 2-MelmpddC, ICG, guanosine (trace impurity from ICG synthesis) and dideoxycytidine monophosphate (ddCMP; a byproduct of 2-MelmpddC hydrolysis), respectively. All peak assignments were performed by comparing the monoanionic m/z of the LC peaks with the expected monoanionic m/z of the proposed species; individual mass spectra that aided the assignments of the LC peaks are omitted for clarity. The ddC-ICG and CMP peaks grow in intensity over the duration of the coincubation experiment, while the 2-MelmpddC peak intensity decreases. X axis: retention time in min; Y-axis: 260 nm UV absorbance in units of mAU (milli-absorbance units).
Figure S16. 2-MelmpddC was coincubated with ICG at pH 8.0. The formation of ddC-ICG dinucleotide was assessed over 7.5 h. The color coding of the LC peaks, and the X- and Y-axes of the stacked liquid chromatograms are the same as Figure S15. All peak assignments were performed by means of mass spectrometry.
Figure S17. (a) MS² spectrum of ddC-ICG dinucleotide. 2-MelmpddC and ICG were coincubated at pH 6.3 for 2 h. The ddC-ICG dinucleotide peak, with a retention time of 19.1 min (Figure 6B, main text), has a major m/z peak of 699.33. This peak was then isolated as the parent ion for the subsequent MS² analysis and partially fragmented. The m/z peaks labeled red arise exclusively from fragments of the 5′-5′-imidazole-bridged ddC-ICG dinucleotide; the m/z peaks labeled blue arise from fragments of the 3′-5′- or 2′-5′-phosphodiester linked ddC-ICG dinucleotide, and the m/z peaks labeled orange could originate from either regioisomer. (b) Possible structures of the ddC-ICG dinucleotide. Left: One of the possible regioisomers of ddC-ICG dinucleotide, with the guanosine and dideoxycytidine nucleotides linked via 3′-5′ or 2′-5′ phosphodiester linkages. Right: The desired regioisomer of the ddC-ICG dinucleotide, with the guanosine and dideoxycytidine nucleotides linked via 5′-5′-imidazole-diphosphate bridge. The bolded arrows indicate the fragmented bond and the associated fragment m/z.
Figure S18. $^{31}$P NMR time course analysis of the 2-MelmpddC-ICG coincubation experiment performed at pH 6.3. $^{31}$P NMR resonances colored red, purple, brown, and green correspond to the resonances of ddC-ICG dinucleotide, 2-MelmpddC, ICG and ddCMP respectively. The formation of ddC-ICG dinucleotide was assessed over 12 h. The 32 to 9 ppm, −1 to −8 ppm and −12 to −26 ppm regions of the $^{31}$P spectra, which have no appreciable peaks, are omitted for clarity. The bottom five traces show the $^{31}$P spectra of the 2-MelmpddC-ICG coincubation experiment at their indicated time points; the top two traces show the result of the 2-MelmpddC degradation experiment (negative control), in which ICG was omitted. All spectra were indirectly referenced to Na$_2$HPO$_4$ in accordance to methods outlined in Section S6.1.2 (Page S34). The negative control experiment clearly showed that the increasing peak at 2.2 ppm was ddCMP, the hydrolysis product of 2-MelmpddC. As expected, the 2-MelmpddC peak at −9.5 ppm decreases over time, due to hydrolysis and reaction with ICG. Throughout the time course, the two new observed peaks at −8.7 and 7.6 ppm are roughly comparable to one another in terms of both peak height and peak integration; the peak intensities of the two new resonances increase from 0 to 4 h, after which they stagnate. Judging from the facts that 1) ddC-ICG dinucleotide is shown to form by means of mass spectrometry, 2) only two new commensurate resonances are observed, and 3) no new resonance is observed in the ppm range wherein phosphodiester resonances typically resides (1 to −5 ppm), we preliminarily conclude that the regiochemistry of the ddC-ICG linkage is via the proposed 5′-5′-imidazole bridge.
Figure S19. 2-MelmpddC was coincubated with PYG at pH 6.3. Possible formation of ddC-PYG conjugate was assessed over a duration of 8.5 h. The purple, brown, and green peaks correspond to 2-MelmpddC, PYG and ddCMP, respectively. All peak assignments were performed by means of mass spectrometry. The elution times of the initially well-resolved PYG and 2-MelmpddC peaks decrease as the coincubation progresses, and the two peaks coelute from the 2.5 h time point and beyond. Regardless of the observed drifts in peak elution time, with the exception of ddCMP, no other peaks are observed to grow in intensity. To confirm that 2-MelmpddC does not react with PYG, the expected monoanionic m/z of the ddCMP-PYG dinucleotide (calculated m/z: 697.14; mass window: ±1 dalton) was used to search the total ion count (TIC) spectra accompanying the shown liquid chromatographs for any trace dinucleotide formation. The focused m/z search reveals that 2-MelmpddC does not react with PYG to any appreciable extent. This indicates that 2-MelmpddC hydrolysis is the only major reaction taking place throughout the 2-MelmpddC-PYG coincubation, and that the nucleophilic 3'-OH, 2'-OH and guanine exocyclic amine groups of PYG does not react appreciably with the activated phosphate of 2-MelmpddC. Thus this experiment provides the first clues that dinucleotide formation between ICG with 2MelmpddN occurs primarily via a 5'-5'-imidazole linkage, and that the regioselectivity for dinucleotide formation is high.
Figure S20. **2-MelmpddC** was coincubated with **PZG** at **pH 6.3**. The formation of ddC-PZG conjugate was assessed over 7.5 h. The red, purple, brown and green peaks correspond to ddC-PZG dinucleotide, 2-MelmpddC, PZG and ddCMP respectively. All peak assignments were performed by means of mass spectrometry. The ddC-PZG dinucleotide peak grows in intensity over the course of the coincubation experiment. Given the results shown in Figure S19 (previous page), we hypothesize the pyrazolyl group of PZG reacts with 2-MelmpddC to form the 5'-5'-pyrazole-bridged ddC-PZG dinucleotide.
6.2 $^{31}$P NMR time course for coincubation experiments with 2-MelmpC & nonhydrolyzable analogues

Dinucleotide formation between 2-MelmpC and various nonhydrolyzable 2-MelmpG analogues were monitored by $^{31}$P NMR for 9 or 10 h. Experimental conditions and methods were identical to those outlined in Section S6.1.2 (page S34), except that 2MelmpddC was replaced by 2-MelmpC. The $^{31}$P resonance of the added Na$_2$HPO$_4$ at pH 6.3 and pH 8.0 was determined to be 0.87 and 2.62 ppm, respectively, using an external reference of 85% H$_3$PO$_4$ (0 ppm by definition) in D$_2$O contained in a coaxial NMR tube. See Figures S21–26 (Page S41–47) for representative data.

![NMR spectra](image)

**Figure S21.** $^{31}$P NMR time course analysis of the 2-MelmpC-ICG coincubation experiment performed at pH 6.3. $^{31}$P NMR resonances colored red, purple, brown, and green correspond to the resonances of rC-ICG dinucleotide, 2-MelmpC, ICG and CMP respectively. The formation of rC-ICG dinucleotide was assessed over 9 h. The $^{31}$P spectral region between 32 to –26 ppm
(Figure S21 cont'd.) was monitored; the 32 to 9 ppm and −9 to −26 ppm regions of the spectra, which have no appreciable peaks throughout the time course analysis, are omitted for clarity. The bottom five traces show the $^{31}$P spectra of the 2-MelmpC-ICG coinubation experiment at their indicated time points. The second top trace shows the reacquired Na$_2$HPO$_4$-referenced $^{31}$P spectra of the 2-MelmpC-ICG sample at the 9 h time point (Na$_2$HPO$_4$ peak was referenced to 0.87 ppm). The top trace shows the Na$_2$HPO$_4$-referenced $^{31}$P trace of the negative control experiment, in which 2-MelmpC and MES buffer were coinubated for 10 h to monitor 2-MelmpC hydrolysis. The top trace clearly shows that the increasing peak at 2.25 ppm is CMP, the hydrolysis product of 2-MelmpC. At pH 6.3, the phosphorus resonance of ICG overlaps significantly with that of CMP. The 2-MelmpC peak at −9.9 ppm decreases over time, due to hydrolysis and reaction with ICG. Throughout the time course, the two new observed peaks at -8.9 and 7.5 ppm are roughly comparable to one another in terms of both peak height and peak integration; their peak intensities increase appreciably from 0 to 3 h, after which they plateau. Judging from the fact that 1) ddC-ICG dinucleotide was shown to form by means of mass spectrometry (See Figure S37, Page S55), and 2) only two new commensurate $^{31}$P resonances are observed, we preliminarily conclude that the formation of the new rC-ICG dinucleotide proceeds with good regioselectivity. As a negative control experiment, PYG was coinubated with 2-MelmpC; no dinucleotide formation is observed (Figure S23–24, Pages S44–45), suggesting that the nucleophilic 3′-OH, 2′-OH and guanine exocyclic amine groups of ICG should not react appreciably with 2-MelmpC. We therefore conclude preliminarily that the regiochemistry of the rC-ICG linkage is via the proposed 5′-5′-imidazole bridge.

The rates of dinucleotide formation and degradation at pH 6.3 were assessed by LCMS analysis. Associated data are shown in Figures S37–41 (Section S6.4, Pages S55–59)
Figure S22. $^{31}$P NMR time course analysis (10 h) of the 2-MelmpC-ICG coinubcation experiment performed at pH 8.0. $^{31}$P NMR resonances colored red, purple, brown, and green correspond to the resonances of rC-ICG dinucleotide, 2-MelmpC, ICG and CMP respectively. At pH 8.0, the resonances of all phosphorus-containing species are baseline-resolved. The spectral regions between 30 to 9 ppm and -9 to -26 ppm are omitted for clarity. The bottom five traces show the $^{31}$P spectra of the 2-MelmpC-ICG coinubcation at their indicated time points. The second top trace shows the reacquired, Na$_2$HPO$_4$-referenced spectrum of the coinubcated sample at the 10 h time point. The top trace shows the Na$_2$HPO$_4$-referenced trace of the 2-MelmpC degradation experiment (negative control) at the 10 h time point (Na$_2$HPO$_4$ peak was determined to be 2.62 ppm). The top trace clearly shows that the increasing peak at 3.8 ppm was CMP. Throughout the time course, the two new observed peaks at -8.7 and 8.4 ppm are roughly comparable to one another both in terms of peak height and peak integration; their peak intensities increase throughout the time course. These two new resonances are attributed to the formation of the 5'-5'-imidazole-bridged CpICG dinucleotide. The rates of CpICG formation and degradation at pH 8.0 are shown in Figures S37–41 (Section S6.4, pages S55–59).
Figure S23. $^{31}$P NMR time course analysis (10 h) of the 2-MelmpC-PYG coincubation experiment performed at pH 6.3. $^{31}$P NMR resonances colored purple, brown, and green are assigned to 2-MelmpC, PYG and CMP respectively. The spectral regions between 32 to 18 ppm and −11 to −26 ppm are omitted for clarity. The bottom five traces show the $^{31}$P spectra of the 2-MelmpC-PYG coincubation at their indicated time points. The second top trace shows the reacquired, Na$_2$HPO$_4$-referenced spectrum of the coincubated sample at the 10 h time point. The top trace shows the Na$_2$HPO$_4$-referenced $^{31}$P trace of the 2-MelmpC degradation experiment (negative control) at the 10 h time point. The increasing peak at 2.2 ppm is CMP; no other substantial phosphorus resonance other that of CMP can be detected throughout the time course. This suggests that PYG does not react with 2-MelmpC to form dinucleotides, and that the potentially nucleophilic groups of PYG, such as the ribose 2'-OH and 3'-OH, and the guanine exocyclic amine moieties, do not react with 2-MelmpC.
Figure S24. $^{31}$P NMR time course analysis (10 h) of the 2-MelmpC-PYG coinubcation experiment performed at pH 8.0. $^{31}$P NMR resonances colored purple, brown, and green are assigned to 2-MelmpC, PYG and CMP respectively. The spectral regions between 30 to 17 ppm and −9 to −28 ppm are omitted for clarity. The bottom five traces show the $^{31}$P spectra of the 2-MelmpC-PYG coinubcation at their indicated time points. The second top trace shows the reacquired Na$_2$HPO$_4$-referenced spectrum of the coinubcated sample at the 10 h time point. The top trace shows the Na$_2$HPO$_4$-referenced $^{31}$P trace of the 2-MelmpC degradation experiment (negative control) at the 10 h time point. The increasing peak at 3.9 ppm is CMP, and no other phosphorus resonance other that of CMP can be detected throughout the time course. This suggests that PYG does not react with 2-MelmpC to form dinucleotides at pH 8.0.
Figure S25. $^{31}$P NMR time course analysis (10 h) of the 2-MelmpC-PZG coincubation experiment performed at pH 6.3. $^{31}$P NMR resonances colored purple, brown, and green are assigned to 2-MelmpC, PZG and CMP respectively. The spectral regions between 32 to 13 ppm and −11 to −26 ppm are omitted for clarity. The bottom five traces show the $^{31}$P spectra of the 2-MelmpC-PZG coincubation at their indicated time points. The second top trace shows the reacquired, $\text{Na}_2\text{HPO}_4$-referenced spectrum of the coincubated sample at the 10 h time point. The top trace shows the $\text{Na}_2\text{HPO}_4$-referenced trace of the 2-MelmpC degradation experiment (negative control) at the 10 h time point. The increasing peak at 2.3 ppm is CMP. Two new peaks are observed at −8.2 and 9.1 ppm, and their peak intensities increase throughout the time course. Given the results shown in Figures S23−24 that the ribose and nucleobase of PZG are unlikely to react with 2-MelmpC, we hypothesized that these two peaks originate from the 5′-5′-pyrazole-bridged CpICG dinucleotide. Despite the fact that PZG can react with 2-MelmpC to yield the 5′-5′-pyrazole bridged dinucleotide, in primer extension assays utilizing PZG as the downstream mononucleotides (Figure 4, main text), PZG does not catalyze primer extension to any significant extent. We hypothesize that the 5′-5′-pyrazole-bridge, which has a protonation $pK_a$ similar to or lower than 1.94 (Figure 1, main text; Figure S2, Page S16), would not be protonated under the examined reaction pH. Thus, nucleophilic attack of the primer 3′-OH group on the pyrazole-bridge would result in the displacement of an anionic pyrazolo-phosphonate group, which is unfavored.
Figure S26. The purified CpICG dinucleotide was subjected to variable pH-\(^{31}\)P NMR analysis to ascertain the possible effect of 5'-5'-imidazole bridge protonation state on the \(^{31}\)P chemical shifts of the two phosphates of CpICG. 100 µL of 10 mM CpICG in D\(_2\)O was prepared, and was referenced with deuterated Na\(_2\)HPO\(_4\); the \(^{31}\)P resonances of Na\(_2\)HPO\(_4\) in this deuterated sample at pH 6.3, 7.0 and 8.0 were determined to be 0.63, 1.38 and 2.61 ppm, respectively, using an external reference of 85% H\(_3\)PO\(_4\) in D\(_2\)O. The deuterated CpICG sample was first adjusted to pD 8, followed by progressive acidification to pD 6.3 with DCl in D\(_2\)O. Throughout the course of sample acidification, the two phosphorus peaks of CpICG shifted downfield (the middle three traces): the upfield \(^{31}\)P resonance of CpICG shifted 0.3 ppm downfield, while the downfield \(^{31}\)P resonance of CpICG shifted 0.8 ppm downfield. The phosphorus resonances of the purified CpICG at pD 8.0 and pD 6.3 are compared with the resonances of CpICG in the corresponding 2-MeImpC-ICG coinubcation experiments (Figure S21–22, pages S41–43; top and bottom traces). The resonance values of CpICG obtained from either source are comparable to one another, with a maximal resonance shift of 0.4 ppm.
6.3 Synthesis, purification and NMR analysis of CpICG

2-MelImpC and ICG were coincubated, in 770 µL scales, with the following components: 75 mM nonhydrolyzable analogues, 75 mM 2-MelImpC, and 200 mM Na•HEPES (pH 8.0). The reaction mixture was adjusted to pH 8.0 using minimal amounts of aqueous HCl or NaOH, followed by incubation at room temperature for 24 h. The reaction mixture was then split into five equal portions and purified by preparative-scale HPLC on a Varian Prostar 210 HPLC system across an Agilent Eclipse XDB C18 column (250 mm × 21.2 mm i.d., 7 µm particle size). The aqueous mobile phase was 25 mM aqueous triethylammonium bicarbonate buffer at ~pH 8.5, and organic mobile phase was acetonitrile. Organic mobile phase was ramped linearly from 2% to 4% over 0-3 min and from 4% to 10% over 3-20 min. The flow rate was maintained at 10 mL min⁻¹. CpICG was eluted from the column at 7.5% CH₃CN in TEAB, which was promptly lyophilized at sub-zero temperature on a VirTis AdVantage 2.0 benchtop lyophilizer (SP Scientific, Warminster, PA) at <50 mTorr over 48 h to afford the titled compound (in triethylammonium cation form) as a white fluffy solid.

¹H and proton-decoupled ¹³C NMR spectroscopies were performed on a Varian Inova 500 MHz spectrometer equipped with a Varian HCN triple resonance cold probe (500 MHz for ¹H, 125 MHz for ¹³C). ¹H-¹H gCOSY, ¹H-¹³C gc2HSQCse and ¹H-¹³C gHMBCAD correlation spectroscopies were performed on an Agilent Inova 700 MHz spectrometer equipped with an Agilent triple resonance helium cold probe (700 MHz for ¹H, 175 MHz for ¹³C). Proton-decoupled ³¹P and ¹H-³¹P gHMBCAD correlation spectroscopies were recorded on a Varian Inova 400 MHz spectrometer equipped with a broadband PFG (z-gradient) probe (400 MHz for ¹H, 161 MHz for ³¹P). Proton and carbon chemical shifts were reported in parts per million (ppm) values on the δ scale. The proton spectrum was referenced to residual protium in D₂O (DHO, δ = 4.79 ppm). The carbon spectrum was referenced to the β carbons of triethylammonium cations (δ = 9.07 ppm). The phosphorus spectrum was referenced to sodium hydrogen phosphate (δ = 0 ppm). All NMR spectra were recorded at 25 °C. Data were reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, q = quartet, m = multiplet, br = broad), and integration.
**1H NMR (D$_2$O)**  
δ 7.90 (s, 1H), 7.87 (d, $J = 2.2$ Hz, 1H), 7.48 (d, $J = 7.6$ Hz, 1H), 5.77 (d, $J = 6.0$ Hz, 1H), 5.75 (d, $J = 2.6$ Hz, 1H), 5.71 (d, $J = 7.6$ Hz, 1H), 4.63 (t, $J = 5.5$ Hz, 1H), 4.34 – 4.31 (m, 1H), 4.19 – 4.15 (m, 1H), 4.13 – 4.03 (m, 4H), 3.99 – 3.94 (m, 3H), 3.11 (TEAH$^+$, q, $J = 7.3$ Hz), 2.46 – 2.42 (d, $J = 1.2$ Hz, 3H), 1.20 (TEAH$^+$, t, $J = 7.3$ Hz).

**13C NMR (D$_2$O)**  
δ 166.73, 161.38, 160.44, 158.16, 155.39, 152.46, 141.93 (dd, $J = 19, 6.4$ Hz), 141.24, 137.21 (dd, $J = 34, 5.5$ Hz), 132.48 (dd, $J = 225, 9.3$ Hz), 117.03, 96.72, 90.54, 87.56, 84.68 (d, $J = 8.7$ Hz), 82.57 (d, $J = 9.2$ Hz), 75.03, 74.66, 71.42, 69.79, 65.26 (d, $J = 5.1$ Hz), 64.57 (d, $J = 4.8$ Hz), 47.45 (TEAH$^+$), 11.31, 9.07 (TEAH$^+$).

**31P NMR (D$_2$O)**  
δ 5.80, −11.02 (pD = 8.0)

**HRMS**  
Calc'd for C$_{20}$H$_{34}$N$_8$O$_7$P [M–H]$^-$: 731.1345; Found: 731.1365

---

**Figure S27.** Full $^1$H NMR spectrum for CpICG.
Figure S28. Full $^{13}$C NMR spectrum for CpICG.

Figure S29. Full $^{31}$P NMR spectrum for CpICG (pH 8.0). The $^{31}$P resonance at 0 ppm is the internal standard, Na$_2$HPO$_4$. 

Figure S30. Full $^1$H–$^1$H gCOSY spectrum for CpICG.

Figure S31. Full $^1$H–$^{13}$C gc2HSQCse spectrum for CpICG.
Figure S32. Full $^1$H–$^{13}$C gHMBCAD spectrum for CpICG.

Figure S33. Full $^1$H–$^{31}$P gHMBCAD spectrum for CpICG.
Figure S34. Full $^1$H NMR spectrum for ICG.

Figure S35. Full $^{13}$C NMR spectrum for ICG.
6.4 Formation and Degradation of purified CplICG

6.4.1 Experimental conditions: liquid chromatography coupled mass spectrometry analyses

CplICG dinucleotide formation from 2-MelmpC and ICG was monitored by the Agilent-ThermoFisher LCMS system (see Section S6.1.1, Page S33) for 10 h. Experimental conditions and methods were identical to those outlined in Section S6.1.1 (page S33), except that 2MelmpddC was replaced by 2-MelmpC. See Figures S37–38 (Page S55–56) for representative data.

Degradation of CplICG at pH 6.3 and 8.0 was monitored by LCMS. Degradation experiments were set up, in 20 μL scales, with 20 mM CplICG and 200 mM buffer (pH 6.3 Na•MES or pH 8.0 Na•HEPES). The sample pH was then promptly adjusted with aqueous hydrochloric acid or sodium hydroxide. At appropriate time points, 2 μL of the 100-fold-diluted samples (400 pmole of CplICG) were injected into the LCMS system. See Figures S39–40 (Pages S57–58) for representative data.

The second-order rate of CplICG formation and first-order rate of CplICG degradation was obtained by monitoring the decay of ICG and CplICG absorbance over time, respectively. See Figure S41 (Page S59) for detailed protocol and data.

Figure S36. Full $^{31}$P NMR spectrum for ICG. The $^{31}$P resonance at 3.8 ppm is the internal standard, trimethyl phosphate.
Figure S37. **2-MelmpC** was co-incubated with **ICG** at **pH 6.3**. The formation of the CplICG dinucleotide was assessed by LCMS over a duration of **10 h**. The red, purple, brown and green LC peaks correspond to CplICG dinucleotide, 2-MelmpddC, ICG and CMP, respectively. All peak assignments were performed by means of mass spectrometry. The CplICG and CMP peaks grow in intensity over the duration of the co-incubation experiment, while the 2-MelmpC peak intensity decreases. The decrease in ICG peak area was used to calculate the second order rate constant for CplICG formation, and relevant data are shown in Figure S41 (Page S59). X axis: retention time in min; Y-axis: 260 nm UV absorbance in units of mAU (milli-absorbance units).
Figure S38. 2-MelmpC was co-incubated with ICG at pH 8.0. The formation of the CpICG dinucleotide was assessed with LCMS over a duration of 10 h. The red, purple, brown and green LC peaks correspond to CpICG dinucleotide, 2-MelmpddC, ICG and CMP respectively. All peak assignments were performed by means of mass spectrometry. At pH 8.0, the CMP peak at the 10 h time point is appreciably smaller than that of pH 6.3, implying that hydrolysis of 2-MelmpC and CpICG is slower at pH 8.0. The decrease in ICG peak area was used to calculate the second order rate constant for CpICG formation, and relevant data are shown in Figure S41 (Page S59).
Figure S39. CpICG was incubated at pH 6.3 under ambient temperature over a duration of 10 h. The degradation of CpICG was monitored by LCMS. The red, brown and green LC peaks correspond to CpICG, ICG and CMP respectively. All peak assignments were performed by means of mass spectrometry. The decrease in CpICG peak area was used to calculate the pseudo-first order rate constant for CpICG hydrolysis, and relevant data are shown in Figure S41 (Page S59).
Figure S40. CplICG was incubated at pH 8.0 under ambient temperature over a duration of 22 h, both in the absence and presence of 50 mM Mg\(^{2+}\). The degradation of CplICG was monitored by LCMS. All peak assignments were performed by means of mass spectrometry. The red LC peak correspond to CplICG. After 22 h, CplICG degrades only minimally, regardless of whether Mg\(^{2+}\) is present (the top and third traces). The brown and green LC peaks correspond to ICG and CMP respectively, and their respective peak areas are less than 5% of that of CplICG. Given the minimal degradation, the pseudo-first order rate of CplICG hydrolysis at pH 8.0 cannot be reliably calculated.
Figure S41. (a) Second order rate plot showing the rate of CpICG formation at pH 6.3 and 8.0, using the LCMS data shown in Figures S37–38 (Pages S55–56). For all the time points taken (from 0 h to 10 h), their respective ICG LC peaks were integrated to afford total absorbance values. Fraction of remaining ICG at each time point was obtained by dividing the integrated ICG absorbance values by that obtained at the 0 h time point, and was used to compute the reciprocal of ICG concentration at each time point (1/\([\text{ICG}]\)). These values were plotted against time and linear regression was then performed. The slope of the fit was divided by 2 to obtain the second order rate \(k_{\text{obs}}\) of CpICG formation. Experiments were performed in duplicate. The \(k_{\text{obs}}\) values of CpICG formation at pH 6.3 and pH 8.0 were estimated to be \((0.9 \pm 0.2) \times 10^{-3} \text{ h}^{-1} \text{ mM}^{-1}\) and \((0.47 \pm 0.07) \times 10^{-3} \text{ h}^{-1} \text{ mM}^{-1}\), respectively. (b) Pseudo-first order rate plot showing the rate of CpICG degradation at pH 6.3, using the LCMS data shown in Figure S39 (Page S57). The fraction of remaining CpICG \((f_{\text{CpICG}})\) at each time point was calculated, and their natural log values \((\ln(f_{\text{CpICG}}))\) were plotted against time. Experiments were performed in duplicate. The estimated \(k_{\text{obs}}\) was 0.063 ± 0.002 h\(^{-1}\).
6.5 Primer extensions with CpICG as activated dinucleotide substrate.

Primer extension reactions were set up, in 10 μL scales, with the following components: 1.2 μM Cyanine 3-labeled primer (Sequence: 5’-Cy3-GACUGACUGG-3’), 2.2 μM complementary template (Sequence: 5’-AACCCGCCAGUCAGUC-3’; bolded nucleotides were not duplexed with primer), 200 mM pH 6.3 Na•MES or pH 8.0 Na•HEPES buffer, 100 mM MgCl₂, and various concentrations of CpICG dinucleotide (Set 1: 1 mM; Set 2: 3 mM; Set 3: 10 mM; Set 4: 20 mM). With this primer-template duplex, CpICG binds to the primer +1 and +2 positions with the activated N-P bond of the 5′-5′-imidazole bridge pointing towards the primer 3′-OH group. For the negative control experiments (Figure 8, main text), the template sequence was modified (Sequence: 5’-AAGCCGCCAGUCAGUC-3’) such that CpICG would be binding to the primer +1 and +2 positions with the unreactive C-P bond of the 5′-5′-imidazole bridge pointing towards the primer 3′-OH group. The detailed procedures for primer extension reactions, as well as subsequent reaction quenching, gel electrophoresis, and data analysis procedures were identical to that outlined in Section S5.2 (Pages S24–25). Time points were taken at 0.25, 0.5, 0.75, 1, 2, 3, 4 and 5 h.

![Diagram](image)

Figure S42. Schematic diagram of the primer extension reaction. The 5′-Cy3-labeled primer was annealed with the template in the presence of 100 mM MgCl₂ and 200 mM buffer of desired pH at room temperature. CpICG was then added to the primer-template complex. The dinucleotide first binds noncovalently to the duplex, followed by an irreversible extension step to afford the primer+1 product.
**Figure S43.** CplICG-driven primer extension reactions at pH 6.3. (a) Top: Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1: Extended primer+C product. Bottom: Plots of the natural log values of the fraction of unreacted primer (ln (P/P₀)) against time. All experiments were performed in triplicate. (b) Saturation curve of the CplICG-driven primer extensions at pH 6.3. The estimated $k_{\text{max}}$ and $K_M$ were $0.28 \pm 0.01$ h⁻¹ and $10.6 \pm 0.8$ mM, respectively ($R^2 = 0.996$).
Figure S44. CplCG-driven primer extension reactions at pH 8.0. (a) Top: Representative 20% denaturing PAGE analyses of the primer extension products. P: Primer; +1: Extended primer+C product. Bottom: Plots of the natural log values of the fraction of unreacted primer (\(\ln (P/P_0)\)) against time. All experiments were performed in triplicate. (b) Saturation curve of the CplCG-driven primer extensions at pH 8.0. The estimated \(k_{\text{max}}\) and \(K_M\) were 0.35 ± 0.02 h\(^{-1}\) and 12 ± 1 mM, respectively \((R^2 = 0.994)\). Thus, pH changes do not appear to alter the \(k_{\text{max}}\) and \(K_M\) of CplCG-driven primer extension to any significant extent (See Figure S43, previous page). See the main text for further discussions on the possible mechanistic origin of this experimental observation.
6.6 \( pK_a \) determination of CplCG.

A 120 \( \mu \)L solution of CplCG in \( \text{D}_2\text{O} \), with 7.77 mM CplCG•2TEAH\(^+\) and 500 mM NaCl, was prepared. The pD of this solution was measured to be 8.66. The acidity of the analyte solution was adjusted by DCI in \( \text{D}_2\text{O} \), followed by proton spectrum acquisition with 128 scans, 1 s relaxation delay, and 45° pulse angle at 25 °C. The chemical shift of the imidazole methine proton of the CplCG 5′-5′-imidazole bridge (highlighted in red, Figure S45a) was traced. Sample acidification and spectral acquisition was repeated until the analyte solution was at pD \( \approx 1.5 \). All proton spectra were referenced to residual protium in \( \text{D}_2\text{O} \) (\( \delta = 4.79 \) ppm). Experiments were performed in duplicate.

Figure S45. (a) Representative stacked spectra for the variable pH-proton spectroscopy of CplCG. Only the aromatic region (7.5–9 ppm) was shown. The methine proton of the 5′-5′-bridged imidazole (labeled red) was traced. As CplCG was acidified beyond pD = 5 (asterisked spectrum), CplCG began to hydrolyze to become CMP and ICG, resulting in additional peaks in the aromatic region. The cytidine base of CplCG also become protonated over the course of pD titration, resulting in the observed downfield shift of the cytidine doublet signal at ~7.6 ppm. (b) The chemical shift values of the imidazole bridge methine proton were plotted against their respective solution pD values, followed by data fitting (Section S3.3, page S14–15) and value correction (Section S3.1, page S13) to estimate the protonation \( pK_a \) of the 5′-5′-imidazole bridge of CplCG. The corrected protonation \( pK_a \) of the CplCG imidazole bridge under \( \text{H}_2\text{O} \) environment is ~5.1.
7. References

(1) Apffel, A.; Chakel, J. A.; Fischer, S.; Lichtenwalter, K.; Hancock, W. S. *Anal. Chem.* **1997**, *69*, 1320.

(2) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059.

(3) Cavalluzzi, M. J.; Borer, P. N. *Nucleic Acids Res.* **2004**, *32*, e13.

(4) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62*, 7512.

(5) Streck, R.; Barnes, A. J. *Spectrochim. Acta A* **1999**, *55*, 1049.

(6) Van Den Berge, E.; Robiette, R. *J. Org. Chem.* **2013**, *78*, 12220.

(7) Kalek, M.; Ziadi, A.; Stawinski, J. *Org. Lett.* **2008**, *10*, 4637.

(8) Matulic-Adamic, J.; Haeberli, P.; Usman, N. *J. Org. Chem.* **1995**, *60*, 2563.

(9) McKenna, C. E.; Higa, M. T.; Cheung, N. H.; McKenna, M.-C. *Tetrahedron Lett.* **1977**, *18*, 155.

(10) Błażewska, K. M. *J. Org. Chem.* **2013**, *79*, 408.

(11) Taylor, S. D.; Mirzaei, F.; Bearne, S. L. *J. Org. Chem.* **2008**, *73*, 1403.

(12) Siddiqui, M. A.; Hughes, S. H.; Boyer, P. L.; Mitsuya, H.; Van, Q. N.; George, C.; Sarafinanos, S. G.; Marquez, V. E. *J. Med. Chem.* **2004**, *47*, 5041.

(13) Izgu, E. C.; Oh, S. S.; Szostak, J. *W. Chem. Commun.* **2016**, *52*, 3684.

(14) Bezençon, J.; Wittwer, M. B.; Cutting, B.; Smieško, M.; Wagner, B.; Kansy, M.; Ernst, B. *J. Pharm. Biomed. Anal.* **2014**, *93*, 147.

(15) Krezel, A.; Bal, W. *J. Inorg. Biochem.* **2004**, *98*, 161.

(16) Rabenstein, D. L.; Sayer, T. L. *Anal. Chem.* **1976**, *48*, 1141.

(17) Gilow, H. M.; Hong, Y. H.; Millirons, P. L.; Snyder, R. C.; Casteel, W. J. *J. Heterocycl. Chem.* **1986**, *23*, 1475.

(18) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N. *Anal. Biochem.* **1989**, *179*, 131.

(19) Connors, K. A. *Binding constants: the measurement of molecular complex stability*; Wiley: New York, 1987.

(20) Schmidtchen, F. P. In *Analytical Methods in Supramolecular Chemistry*; Schalley, C., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Germany, 2007; pp 55–78.

(21) Tam, C. P.; Fahrenbach, A. C.; Björkbom, A.; Prywes, N.; Izgu, E. C.; Szostak, J. W. J. *Am. Chem. Soc.* **2017**, *139*, 571.

(22) Thordarson, P. *Chem. Soc. Rev.* **2011**, *40*, 1305.

(23) Joyce, G. F.; Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1984**, *176*, 279.

(24) Rio, D. C.; Ares Jr., M.; Hannon, G. J.; Nilsen, T. W. *Cold Spring Harb. Protoc.* **2010**, *2010*, pdb. prot5444.