This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

Cite This: J. Am. Chem. Soc. 2018, 140, 783−792

pubs.acs.org/JACS

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, United States

[‡]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

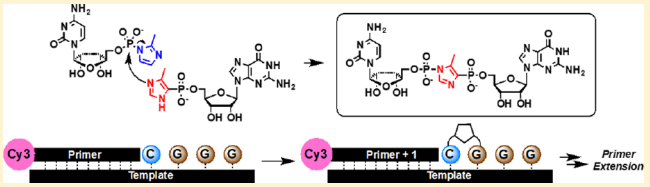
[§]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, United States

[⊥]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, United States

Supporting Information

ABSTRACT: We report the synthesis of guanosine 5′-(4-methylimidazolyl)phosphonate (ICG), the third member of a series of nonhydrolyzable nucleoside 5′-phosphoro-2-methylimidazolide (2-MeImpN) analogues designed for mechanistic studies of nonenzymatic RNA primer extension. The addition of a 2-MeImpN monomer to a primer is catalyzed by the presence of a downstream activated monomer, yet the three nonhydrolyzable analogues do not show catalytic effects under standard mildly basic primer extension conditions. Surprisingly, ICG, which has a pK_a similar to that of 2-MeImpG, is a modest catalyst of nonenzymatic primer extension at acidic pH. Here we show that ICG reacts with 2-MeImpC to form a stable 5′−5′-imidazole-bridged guanosine-cytosine dinucleotide, with both a labile nitrogen−phosphorus and a stable carbon−phosphorus linkage flanking the central imidazole bridge. Cognate RNA primer−template complexes react with this GC-dinucleotide by attack of the primer 3′-hydroxyl on the activated N−P side of the 5′-5′-imidazole bridge. These observations support the hypothesis that 5′−5′-imidazole-bridged dinucleotides can bind to cognate RNA primer−template duplexes and adopt appropriate conformations for subsequent phosphodiester bond formation, consistent with our recent mechanistic proposal that the formation of activated 5′−5′-imidazolium-bridged dinucleotides is responsible for 2-MeImpN-driven primer extension.



INTRODUCTION

For the earliest cellular organisms,[1,2] replication must have occurred through primitive mechanisms without elaborate enzymatic machinery.[3,4] Template-directed nonenzymatic primer extension[5] is an important experimental model system for the study of chemically driven genome replication. In this model, chemically activated mononucleotides or short oligonucleotides[6] bind to a template,[7] followed by monomer polymerization[8] or oligomer ligation[9] to assemble the template complement. Imidazoles, among other heteroarenes,[10,11] have been broadly utilized in the chemical activation of mono-[12,13] and oligonucleotides.[6,14] Much effort has been invested in enabling mixed sequences, especially those containing adenosine and uridine,[15] to be copied efficiently and with high fidelity. Among these efforts, a small, focused screen of substituted imidazoles as monomer leaving groups revealed 2-aminoimidazole-activated monoribonucleotides as superior primer extension substrates compared to imidazole- and 2-methylimidazole-activated monomers.[14] On the other hand, activated trinucleotides that bind downstream of the extending nucleotide were discovered to be much better catalysts than activated monomers, not only providing rate enhancements of at least 2 orders of magnitude, but also enabling copying of templates containing all four canonical nucleotides in one pot.[6] Additional work has been performed to determine and enhance the fidelity,[16,17] regioselectivity,[18−20] and thermodynamics[21−23] of these reactions. Plausible prebiotic conditions conducive to imidazole synthesis have also been probed.[24,25]

For decades, the classical S_N2-type mechanism was thought to constitute the sole mechanistic pathway for primer extension by nucleotide phosphoroimidazolides:[3] the nucleophilic 3′-hydroxyl group of the primer would attack the imidazole-activated phosphate of the incoming monomer via an in-line mechanism, and concomitantly displace the imidazole leaving group. Activated monomers[26] and oligomers[6,14] binding downstream of the polymerization site play important catalytic roles, as their presence strongly

Received: November 6, 2017
Published: December 18, 2017

© 2017 American Chemical Society

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792

accelerates the rate of primer extension. This rate acceleration was initially hypothesized to result from noncovalent leaving group−leaving group interactions between consecutively bound activated substrates, which preorganize the upstream monomer and lower the activation barrier for subsequent phosphodiester bond formation. However, emerging evidence suggests that two activated monomers can react with each other to form a highly activated imidazolium-bridged dinucleotide intermediate (Np-MeIm-pN, Figure 1), and the

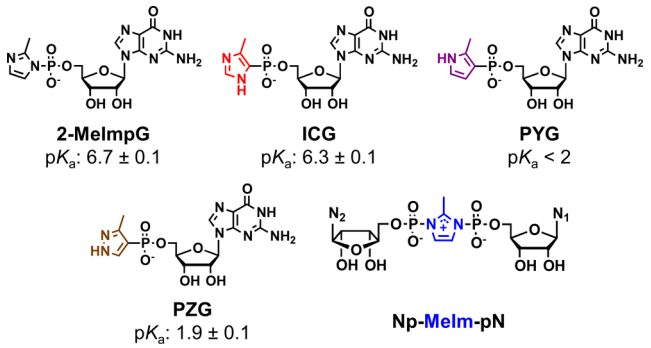


Figure 1. Chemical structures of 2-MeImpG, ICG, PYG, PZG, and Np-MeIm-pN, together with the pK_a values of the heteroaryl leaving groups in water, as estimated using NMR spectroscopy (mean and range of duplicate measurements). The pK_a of the pyrrolyl group of PYG (purple) could not be confidently determined due to extensive proton−deuterium exchange at strongly acidic pHs (Figure S3, Supporting Information).

observed rate acceleration afforded by a downstream activated monomer (or oligomer) may be due to covalent nucleophilic catalysis.[27] Hence, we have sought to ascertain which of these two potential pathways is primarily responsible for catalysis by downstream activated nucleotides.

Nonhydrolyzable nucleotide analogues have been indispensable tools in crystallographic and mechanistic studies of polymerases and ligases.[28] For example, nonhydrolyzable nucleotide triphosphates, such as α,β-methylene[29] and α,β-difluoromethylene[30] nucleoside triphosphates have been cocrystallized with T7 RNA polymerase[29] and DNA polymerase β,[30] respectively. Also, Raines and co-workers used α,β-methylene GTP to show that the scission of the GTP $\alpha-\beta$-phosphoanhydride bond is critical for RNA ligation by the archaeal RNA ligase RtcB.[31] Drawing inspiration from these strategies, we have developed several nonhydrolyzable 2-MeImpN analogues as mechanistic probes of nonenzymatic RNA template copying. We previously reported the synthesis of guanosine 5′-(3-methylpyrazolyl) phosphonate (PZG, Figure 1)[32] and guanosine 5′-(2-methylpyrrolyl)phosphonate (PYG, Figure 1)[14] as nonhydrolyzable analogues of 2-MeImpG. Our crystal structures of RNA-PZG complexes revealed[32] that the noncovalent interactions between monomers and RNA templates do not always obey canonical Watson−Crick geometry under crystallization conditions. More importantly, we were not able to observe any noncovalent interactions between the leaving groups of consecutively bound PZG nucleotides. We suspected that the low pK_a of the pyrazolyl group of PZG might have led to chemical differences from 2-MeImpG that prevented the formation of catalytically relevant interactions in the crystal structure. In a parallel approach, we used P^1,P^3-diguanosine-5′-triphosphate (GpppG) as a stable analogue of the corresponding activated Gp-MeIm-pG dinucleotide for crystallographic studies, and showed that it could have the appropriate geometry to bind to cognate RNA primer−template complexes via Watson−Crick base pairing, with its 5′−5′-bridge preorganized for subsequent S_N2 attack by the primer 3′-hydroxyl group.[33]

Here we report the synthesis of guanosine 5′-(4-methylimidazolyl)phosphonate (ICG, Figure 1), the third and closest nonhydrolyzable isosteric analogue of 2-MeImpG. For all three analogues (ICG, PYG, and PZG), we report the pK_a of the leaving group mimics in the context of a full nucleotide, along with the binding constants of 2-MeImpG and all three pseudo-activated nucleotides for RNA primer−template complexes. Finally we report measurements of the pH-dependent catalytic activity of all four nucleotides in nonenzymatic primer extension. None of the three analogues shows catalytic effects under mild alkaline primer extension conditions, yet ICG, which has a pK_a and RNA affinity similar to those of 2-MeImpG, is a modest catalyst under acidic pH. Using NMR and mass spectrometry, we demonstrate that ICG reacts with an activated C monomer to form the 5′−5′-imidazole-bridged guanosine-cytosine dinucleotide (CpICG) with very high regioselectivity. The purified GC-dinucleotide, with a labile nitrogen−phosphorus and a stable carbon−phosphorus linkage flanking the imidazole bridge, reacts with cognate RNA primer−template complexes to afford extended primers only when the labile nitrogen−phosphorus linkage, but not the stable carbon−phosphorus linkage, is proximal to the primer 3′-hydroxyl group. These results are consistent with our recent mechanistic proposal that 2-MeImpN-driven primer extension proceeds via the formation of 5′−5′-imidazolium-bridged dinucleotide intermediates.

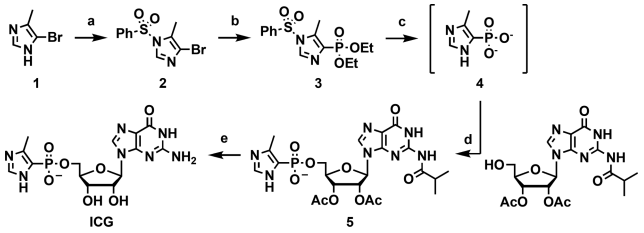
RESULTS

Synthesis and pK_a Determination of Guanosine 5′-(4-Methylimidazolyl)phosphonate (ICG). Our synthetic strategy for the construction of ICG was parallel to our previously described synthesis of PZG.[32] Briefly, the commercially available 4-bromo-5-methylimidazole **1** was protected on N3 with a benzenesulfonyl group to give **2**,[34] followed by installation of the phosphonate moiety via a palladium-catalyzed carbon−phosphorus coupling[35] reaction to afford **3**. The protected C−P-linked imidazolylphosphonate was then coupled to the 5′-hydroxyl of guanosine to afford **5**, by first removing the phosphonate alkyl groups (**4**),[36] followed by Mitsunobu phosphonylation.[37] Subsequent deprotection of **5** then furnished the desired ICG product (Scheme 1).

The pK_a of the leaving group of monomers in the downstream (i.e., primer+2) position has been shown to strongly influence the rate of reaction of the primer with an activated monomer in the adjacent (i.e., primer+1) position,[14] suggesting that the protonation state of the downstream monomer plays an important role in catalysis. Hence, we sought to determine the protonation pK_a of the leaving group of 2-MeImpG and the heteroaryl-phosphonate moieties of nonhydrolyzable 2-MeImpG analogues in the context of the full nucleotide (Figure 2A), and to correlate the experimentally determined pK_a values with catalysis under primer extension conditions. We used proton NMR spectroscopy for pK_a determination[38,39] by monitoring the chemical shifts of the aromatic protons of 2-MeIm and its mimics as a function of pD (Figure 2B). The resulting sigmoidal curves (Figure 2C) were fit to an equation derived from the Henderson−Hasselbach

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792

Scheme 1. Synthesis of ICG[a]



[a]Reaction conditions: (a) $PhSO_2Cl$, Et_3N, DCM, 48%; (b) $(EtO)_2P(O)H$, $Pd(PPh_3)_4$, PPh_3, Et_3N, DMSO, 115 °C, 2 h, 35%; (c) TMSBr, Et_3N, DCM, 4 h; then MeOH; (d) N^2-isobutyryl-2′,3′-diacetylguanosine, DIAD, PPh_3, DCM, 3 h; (e) conc NH_3 in H_2O, MeOH, 65 °C, 4 h (22% over c–e). See Figures S34–S36 (SI) for the ^1H, ^{13}C, and ^{31}P NMR spectra of ICG.

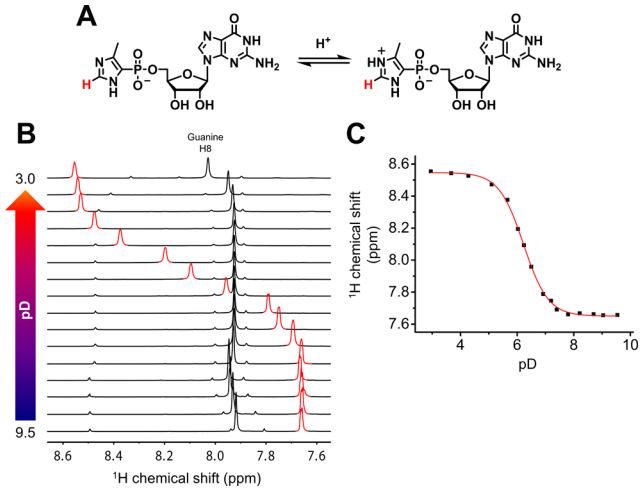
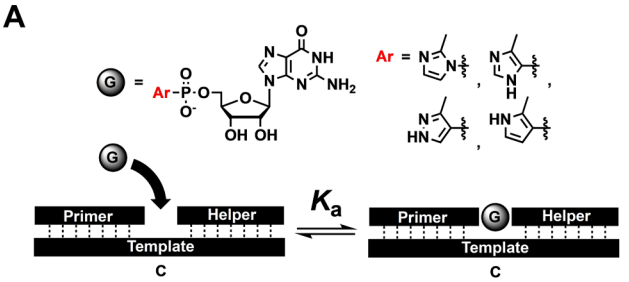


Figure 2. (A) Protonation of the imidazolylphosphonate of ICG. The chemical shift of the ICG imidazole methine proton (colored red) is monitored throughout the pD titration experiment. (B) The stacked variable pD-proton NMR spectra of ICG. The aromatic region of the proton spectra (7.5–8.7 ppm) is shown. The peaks in the stacked proton spectra corresponding to the ICG imidazole methine proton are colored red. All spectra are referenced to the residual HOD peak at 4.79 ppm (not shown). As the monomer solution was acidified, the methine proton shifted downfield dramatically, crossing over the proton peak of the guanine C8 proton. Experiments were performed in duplicate. (C) The corresponding chemical shift vs monomer pD plot. The inflection point of the sigmoidal fit curve reflects the pK_a value of the leaving group moieties of monomer or analogue.

formula (Section S3.3, Supporting Information (SI)), followed by correction as reported by Krezel and Bal[40] (Section S3.1, SI) to estimate the pK_a values of the four monomers (Figure 1). Of the nonhydrolyzable analogues, the protonation pK_a of the imidazolyl moiety of ICG (pK_a = 6.3) is most similar to that of the 2-MeIm leaving group of 2-MeImpG (pK_a = 6.7, literature value[41] = 7.09; Figure S1, SI). The pK_a values of the heteroaryl groups of PZG and PYG are at least 4 units lower than that of ICG (Figures S2 and S3, SI). It is interesting to note that the pK_a values of ICG and 2-MeImpG are at least 1 unit lower than the reported pK_a (7.86) of free 2-methylimidazole[42] despite the expected stabilization of the protonated state by the adjacent negative charge of the phosphate or phosphonate. Apparently, the inductive electron withdrawing effect of phosphate or phosphonate groups has a greater influence and serves to lower the pK_a of the leaving groups relative to free 2-methylimidazole.

Affinity of 2-MeImpG and Nonhydrolyzable Analogues to RNA Duplex. We assessed the affinity of 2-MeImpG and its nonhydrolyzable analogues for cognate binding sites on an RNA primer–template complex by isothermal titration calorimetry.[43,44] Specifically, 4.4 equiv of 2-MeImpG or analogue was titrated into 1.5 mM of a previously reported interrupted RNA duplex in which the binding site for guanosine mononucleotides is sandwiched between the primer 3′-terminus and a downstream helper oligonucleotide, in the presence of pH 7-buffered NaCl solution (500 mM Na^+) at 20 °C (Figure 3A). The helper



	K_a (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
2-MeImpG	(2.1±0.4)×10³	−4.45±0.12	−17.1±3.0	−43.1±10.7
ICG	(4.1±0.5)×10³	−4.84±0.07	−14.1±0.1	−31.5±0.5
ICG with Mg^{2+}	(4.5±1.0)×10³	−4.89±0.13	−15.2±0.5	−35.1±1.9
PZG	(2.8±0.3)×10³	−4.63±0.06	−14.2±0.6	−32.7±1.8
PYG	(3.6±0.2)×10³	−4.78±0.03	−14.8±0.3	−34.2±0.9

Figure 3. Affinity of 2-MeImpG and nonhydrolyzable analogues with the primer–template–helper (P/T_S/H) sandwich duplex. (A) The binding scheme between 2-MeImpG and its analogues with the P/T_S/H duplex. The analogue is sandwiched between two flanking G residues at the primer 3′-terminus and the helper 5′-terminus, respectively. (B) Associated thermodynamic parameters, ±standard error, n = 3. See Figure S4 for the LC-HRMS analysis of the sandwich oligonucleotides; see Figure S5 and Table S1 (SI) for data derived from individual ITC titrations. All ITC experiments were performed at 20 °C.

oligonucleotide primarily serves to enhance the affinity of monomer or analogue to the RNA primer–template complex, thus enabling affinity be measured by ITC.[21] We now show that the affinities of the nonhydrolyzable 2-MeImpG analogues for the sandwich duplex are comparable, with estimated K_a values between 2.8 × 10^3 and 4.5 × 10^3 M^{-1}; this is slightly higher than the K_a of 2-MeImpG ((2.1 ± 0.4) × 10^3 M^{-1}, Figure 3B). The estimated stoichiometric constant n of RNA:analogue binding for all three monomer analogues was ~0.8, consistent with 1:1 analogue-to-RNA binding. The free energy of RNA binding for the analogues ranged between −4.6 and −4.9 kcal mol^{-1}, suggesting that the chemical characteristics of the heteroaryl groups of the analogues have only minimal effects on the free energy of binding with RNA. Notably, the presence of 50 mM Mg^{2+} did not alter the affinity of ICG for RNA. We suspect that the high ionic strength of both solutions negates any effect Mg^{2+} may have on the binding of the analogue to RNA duplexes.

Effects of 2-MeImpG and Nonhydrolyzable Analogues on RNA Primer Extension. With these insights

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783–792

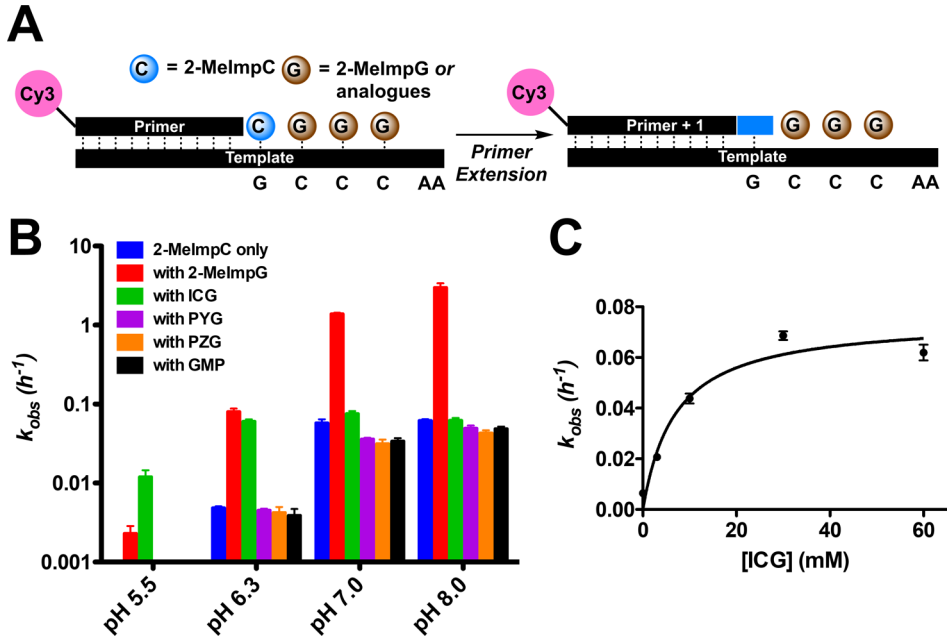


Figure 4. (A) Schematic of the primer extension assays used herein. 2-MeImpC binds to the cognate primer +1 binding site, followed by the consecutive binding of multiple 2-MeImpG or its nonhydrolyzable analogues. (B) k_{obs} of the primer extension shown in (A) under different reaction pH and different stacking guanosine monomers. (C) Saturation curve of the ICG-assisted primer extensions at pH 6.3. Monomer precipitation was not observed, even at the highest concentration of ICG (60 mM). In (B) and (C), all primer extensions were performed in triplicate. See Figures S7–S14 (SI) for representative denaturing PAGE gel images and first-order rate plots for the primer extensions depicted in (B), as well as the saturation curves of 2-MeImpG- and ICG-catalyzed primer extensions at pH 6.3 and 8.0.

into the chemical and thermodynamic properties of the 2-MeImpG analogues, we turned to the question of catalysis of the reaction between the primer and an adjacent monomer, by analogues bound to the template at the primer +2 position. We measured the pseudo-first-order rate of conversion of primer to extended products on a template with a GCCCAA-5′ overhang (Figure 4A; Figure S6, SI), using a reaction mixture that contained a large excess of the 2-MeImpC (30 mM). 2-MeImpG and its nonhydrolyzable analogues were used at 30 mM, because higher concentrations led to observable precipitation, especially for ICG and GMP; the catalytic metal ion Mg^{2+} was present at 100 mM. We varied the pH of the primer extension reactions from mildly acidic (pH 5.5) to slightly alkaline (pH 8) to probe the effect of pH on catalysis by the analogues. Interestingly, the rate of primer extension with 2-MeImpC in the absence of any downstream guanosine nucleotides (blue bars, Figure 4B) was slightly higher than in the presence of unactivated GMP (black bars, Figure 4B). We attribute this small effect to the potential for weak binding of 2-MeImpC at the primer +2 position in the absence of a competing G nucleotide. To avoid this confounding effect, we compare the catalytic effects of 2-MeImpG and the three analogues to the baseline rate observed in the presence of GMP. As expected 2-MeImpG in the primer +2 position exhibits a strong catalytic effect at all pH values from 5.5 to 8.0 (red bars, Figure 4B). In general, all primer extension reactions exhibit higher rates at alkaline pH than in neutral or acidic pHs, presumably due to the need to deprotonate the primer 3′-hydroxyl. We then asked whether binding of ICG, PYG or PZG at the primer +2 position could enhance the rate of primer extension with 2-MeImpC (green, purple, and orange bars, respectively; Figure 4B). Regardless of pH, the rate of primer extension was not improved in the presence of PYG or PZG, compared with GMP. Therefore, neither PZG nor PYG affords any catalysis to primer extension.

On the other hand, ICG, which has an imidazolyl pK_a close to the pK_a of 2-MeImpG, did exhibit a pH-dependent catalytic effect on primer extension (Figure 4B). At a primer extension pH of 8, the presence of ICG led to only a very small increase in the rate of primer extension of 20% (compared to the 60-fold effect of 2-MeImpG). However, when the extension reactions were carried out at neutral pH, the catalytic effect of ICG increased to ~2.5-fold (vs 40-fold for 2-MeImpG), while at pH 6.3, ICG exhibited a 15-fold rate enhancement relative to GMP, rivaling the ~20-fold effect of 2-MeImpG. The k_{obs} of ICG-assisted primer extension is similar from pH 8.0 to 6.3, while the k_{obs} of 2-MeImpG-assisted primer extension decreased ~37 fold from pH 8.0 to pH 6.3. At pH 5.5, primer extension in the presence of GMP could barely be detected even after 20 h. Even 2-MeImpG in the primer +2 position yielded less than 5% extended primer after 20 h. However, the catalytic effect of ICG surpassed that of 2-MeImpG and resulted in readily observable primer extension, leading to 20% of the primer +1 product after 20 h (Figure S7, SI). Thus, the rate enhancement afforded by ICG is much more pronounced at acidic pH, while the rate enhancement afforded by 2-MeImpG is most pronounced at alkaline pH.

To see if the increasing catalytic effect of ICG at lower pH (and conversely, the declining catalytic effect of 2-MeImpG at lower pH) was due to changes in template binding, we measured the catalytic effect as a function of concentration at pH 6.3 and 8.0. Primer extensions at pH 6.3 were repeated at four different concentrations of ICG (0, 3, 10, 30, 60 mM). The estimated k_{max} was 0.075 ± 0.008 h^{-1}, and the K_M was 7 ± 3 mM. From the rate vs concentration plot (Figure 4C), it is apparent that the rate-enhancement afforded by ICG approaches a maximum when the concentration of ICG is at,

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783–792

or above 20 mM. Since both ICG and 2-MeImpG achieve their maximal catalytic effect at 30 mM at pH 6.3 (Figure S13, SI), their different catalytic abilities cannot be attributed to differential binding. In contrast, at pH 8.0, the rates of primer extension are unaffected by ICG concentration (0 mM ICG; rate ≈ 0.07 h^{-1}; Figure S12, SI). Since at pH 8 ICG and 2-MeImpG bind with similar affinities to primer/template complexes, the pH-dependence of catalysis by ICG does not reflect pH-dependent changes in binding, but likely reflects a direct pH-dependence of the catalytic mechanism.

How can we explain the observation that ICG becomes a better catalyst of primer extension at progressively lower pH, while 2-MeImpG gradually loses catalytic ability at low pH? Given the observation of a 5′−5′-imidazolium-bridged dinucleotide in primer extension reactions, we wondered if ICG might catalyze primer extension in a similar fashion, with the imidazolyl group of ICG (highlighted in blue, Figure 5)

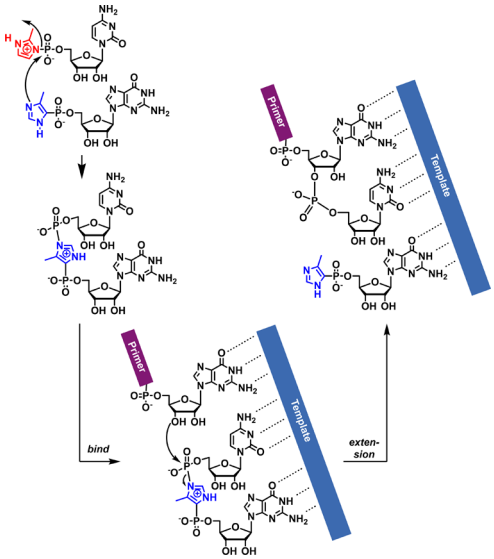


Figure 5. Proposed mechanism of the ICG-assisted primer extension via the formation of the 5′−5′-imidazole-bridged dinucleotide (CpICG).

reacting with 2-MeImpG to form a mixed N−P- and C−P-linked imidazole-bridged intermediate (Figure 5, middle panel).[27] Due to the lower pK_a of ICG than 2-MeImpG, the ratio of protonated 2-MeImpG to unprotonated ICG is increased at lower pH, while a higher proportion of ICG remains in the neutral, nucleophilic form. Therefore, we sought to examine whether the proposed 5′−5′-imidazole-bridged dinucleotide was formed during the ICG-catalyzed primer extension.

Detection of 5′−5′-Imidazole-Bridged Mixed-Linkage Dinucleotide via LCMS. Initial attempts to identify the 5′−5′-imidazole-bridged mixed-linkage dinucleotide were performed with analytical reversed-phase liquid chromatography-coupled mass spectrometry (LCMS), using ICG and dideoxycytidine 5′-phosphoro-2-methylimidazolide (2MeImpddC) as the substrate. 2MeImpddC was chosen in lieu of 2-MeImpC because its dideoxyribose moiety affords enhanced hydrophobicity, allowing 2MeImpddC, as well as its ICG-conjugated products to be chromatographically resolved from ICG. Additionally, removal of the cytidine ribose 2′- and 3′-hydroxyl groups eliminates the possibility of 2′-5′- and 3′-5′-linked CC dinucleotide formation, thus facilitating chromatographic separation of possible products. 2MeImpddC was conveniently prepared from Zalcitabine with a three-step protection−activation−deprotection sequence (Figure 6A). First, 30 mM ICG and 30 mM 2MeImpddC were incubated at pH 6.3 in the absence of Mg^{2+}. After 2 h, the incubated sample was chromatographically analyzed under reverse phase conditions (300 pmol each for ICG and 2MeImpddC). ICG and 2MeImpddC were well resolved, and only one new peak appeared, with a retention time of 19.11 min (Figure 6B) and an m/z of 699.33 (Figure 6C), which corresponds to the expected m/z of monoanionic ddCMP-ICG dinucleotide. To confirm that the ddCMP-ICG dinucleotide carries the desired 5′−5′-imidazole bridge, we performed tandem MS fragmentation on the precursor ion with the m/z = 699.33 (with a m/z window of ±1). The fragmentation voltage was chosen such that the precursor ion was only partially fragmented so that the major fragment m/z peak could indicate which bond within the precursor ion is the most labile (Figure 6D). Notably, the major fragment peak, with an m/z of 416.17 (peak 1, Figure 6D), is suggestive of fragmentation of the 5′−5′-imidazole-bridged ddCMP-ICG dinucleotide, with heterolytic cleavage occurring at the O−P bond between the guanosine and the imidazole bridge moieties (Figure 6E). Other possible regioisomers of ddCMP-ICG dinucleotide, such as the 3′-5′-, 2′−5′-, and nucleobase-linked dinucleotide, could not afford fragments with m/z of 416.17. This observation thus provides the first clue that ICG could react with activated monomers to afford 5′−5′-imidazole-bridged dinucleotides.

Structure Elucidation of CpICG by NMR. Encouraged by the LC-MS results, we then asked whether 2-MeImpC and ICG could react to form the 5′−5′-imidazole-bridged, mixed linkage CG dinucleotide (CpICG, Figure 5, middle panel). 50 mM ICG was incubated at pH 6.3 with 50 mM 2-MeImpC in the absence of Mg^{2+}, and the progress of dinucleotide formation was monitored by ^{31}P NMR spectroscopy overnight (Figure 7A). As seen in the t = 0 h spectrum, acquired immediately after mixing, two major ^{31}P singlet peaks were seen at −9.89 and 2.30 ppm, corresponding to 2-MeImpC and ICG, respectively. Three additional minor singlet peaks were observed at −8.95, 2.25, and 7.54 ppm, and no other significant ^{31}P peaks are observed. These peaks increased in intensity as the incubation proceeded, with the peaks at −8.95 and 7.54 ppm maintaining similar intensity throughout the entire NMR time course, suggesting that they might arise from the same molecule. The peak at 2.25 ppm is assigned to CMP by internal referencing, which results from the hydrolysis of 2-MeImpC (Figure S21, SI). The remaining two ^{31}P peaks at −8.95 and 7.54 ppm were tentatively assigned to the two phosphorus atoms of the newly formed GC dinucleotide. To establish the regiochemistry of the newly formed GC dinucleotide, and rule out the possibility of 2′-5′-, 3′-5′-, or guanine-phosphate linkages, we examined the reaction of PYG with 2-MeImpC. The pyrrolylphosphonate group of PYG is known to be non-nucleophilic under near-neutral conditions. Throughout an overnight ^{31}P NMR time course, the only new detectable product was CMP (Figure 7B), suggesting that the nucleophilic ribose hydroxyls and nucleobase exocyclic amine of PYG do not react with 2-MeImpC over the time course of primer extension assays. Since the only reactive site on ICG is the imidazolylphosphonate group, the CpICG dinucleotide should indeed contain a 5′−5′-imidazole bridge.

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792

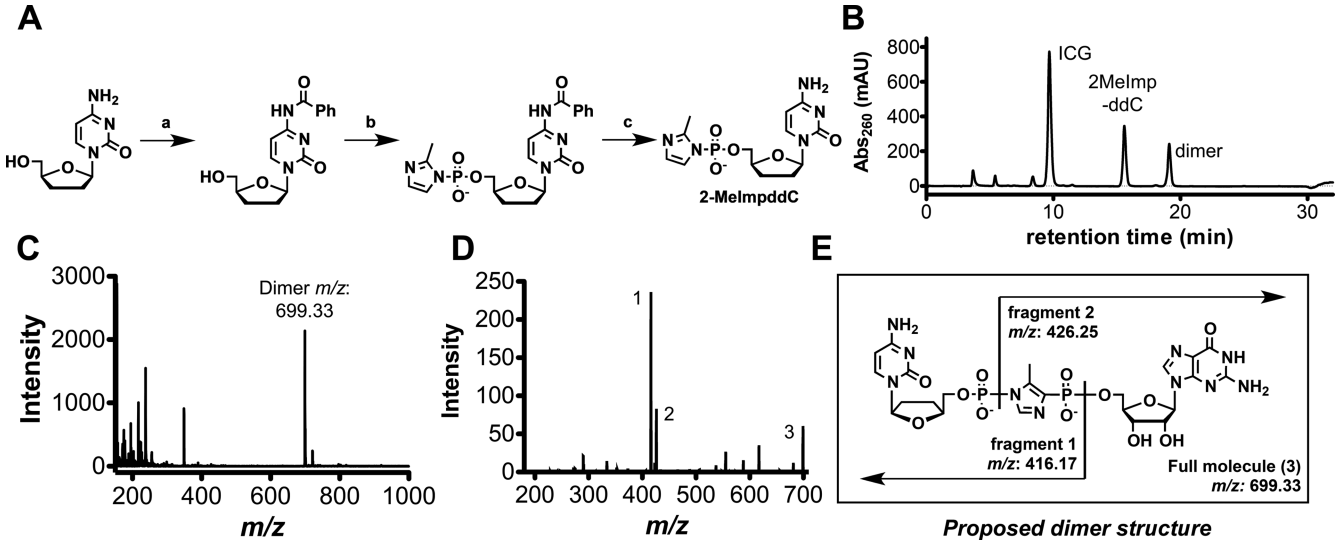


Figure 6. (A) Synthesis of 2-MeImpddC. Conditions: (a) TMSCl, pyr; then BzCl; then conc NH_4OH; (b) $POCl_3$, DIPEA, $PO(OMe)_3$; then 2-MeIm; then 1 M TEAB; (c) conc NH_4OH, 65 °C. (B) Liquid chromatography analysis of the products of 2MeImpddC-ICG coincubation at pH 6.3 after 2 h. UV absorbance of the products at 260 nm was monitored. (C) ESI-MS analysis of the LC peak at ∼19 min shown in (B). The major m/z peak corresponds to the monoisotopic monoanionic mass of the proposed ddCMP-ICG dinucleotide; the structure of the proposed dinucleotide is shown in (E). (D) The MS^2 spectrum obtained by fragmentation of the m/z = 699.33 precursor ion. Peak 3 is the partially fragmented parent ion; peak 2, with a m/z of 426.25, corresponds to the ICG fragment of CpICG, resultant from P−N bond cleavage during collision-induced dissociation. See Figure S17 of SI for a more detailed assignment of the MS^2 spectrum. (E) The proposed structure of the 5′−5′-imidazole-linked ddCMP-ICG dinucleotide, the observed m/z signals, and the corresponding CpICG fragments. 2-MeImpddC had been coincubated with ICG, PYG, and PZG, and the formation of 5′−5′-heteroarene-bridged dinucleotide had been monitored by LCMS and NMR. See Figures S15−S20 (SI) for detailed analyses.

Because ICG catalyzed primer extension at pH 6.3 but not at pH 8.0, we asked whether the 5′−5′-imidazole-bridged CpICG dinucleotide could form under basic conditions. Using ^{31}P NMR spectroscopy, we monitored the overnight reaction of coincubated ICG and 2-MeImpC at pH 8.0 in the absence of Mg^{2+}. As with the pH 6.3 NMR time course, three new singlet peaks increased in intensity over time. The three new peaks were shifted downfield to varying degrees, presumably due to the change in pH (Figure 7C). The peak at 3.85 ppm was assigned to CMP (Figure S22, SI) suggesting that the remaining two peaks belong to CpICG. This was confirmed with the same PYG control experiment as used above (Figure 7D). To confirm that the GC dinucleotide formed at pH 6.3 and 8.0 is indeed the same chemical entity, the dinucleotide prepared from overnight 2-MeImpC-ICG coincubation at pH 8.0 was chromatographically purified under reverse-phase conditions, followed by lyophilization and ^{31}P NMR acquisition at pH 6.3. The two singlet ^{31}P peaks thus obtained have chemical shift values close to those obtained in the pH 6.3 ^{31}P NMR time course (within ∼±0.2 ppm; Figure S26, SI), thus confirming that the change in dinucleotide phosphorus chemical shifts is due to pH change.

The proposed structure of CpICG, as well as the presence of the 5′−5′-imidazole-bridged connectivity, was validated by a combination of 1D homonuclear and 2D heteronuclear coupled NMR spectroscopy (Figure 8A,B). We looked specifically for the additional ^{31}P-induced splitting of the carbon resonances associated with the bridging imidazole of CpICG. To identify these carbon resonances, we used both $^{1}H-^{13}C$ coupled gc2HSQCse and gHMBCAD spectroscopy. The doublet ^{1}H peak at 7.87 ppm was assigned to the imidazolyl methine proton, while the ^{1}H peak at 2.44 ppm was assigned to the methyl protons exocyclic to the imidazole bridge (Figure S27, SI). As shown in the partial $^{1}H-^{13}C$ coupled gc2HSQCse spectrum, the imidazolyl methine proton is correlated with the doublet of doublet resonance at 141.8−142.0 ppm (Figure 8C,F). This is consistent with the expected splitting pattern induced by both the N−P and C−P phosphorus atoms. On the other hand, the partial $^{1}H-^{13}C$ coupled gHMBCAD spectrum clearly demonstrates that the exocyclic methyl protons of the bridging imidazole are coupled to only two carbon resonances, which must be the imidazole bridge carbons that are two and three bonds away (Figure 8D,F). These resonances at 137.0−137.4 and 131.5−133.4 ppm were also doublets of doublets, consistent with the 5′−5′-phosphate-imidazole-phosphonate linkage regiochemistry. The corresponding carbon signals of ICG only have doublet splitting because the molecule only contains one phosphorus atom (Figure S35, SI). Finally, $^{1}H-^{31}P$ coupled gHMBCAD spectroscopy of CpICG (estimated $^{n>1}J_{P,H}$ = 8 Hz), clearly showed that the three protons of the imidazole exocyclic methyl group are coupled to both phosphorus atoms of the imidazole-flanking phosphate and phosphonate (Figure 8E). These data, together with the ^{31}P NMR time courses shown in Figure 7, strongly suggest that ICG reacts with 2-MeImpC to form the 5′−5′-imidazole-bridged CpICG.

Primer Extension Reactions with CpICG. Having established that 5′−5′-linked CpICG can form by the reaction of ICG with 2-MeImpC, we asked whether this dinucleotide can bind to its cognate RNA template and react with an upstream primer. We carried out the primer extension experiment as in Figure 4, using the same RNA primer−template complex and reaction conditions as used for the pseudo-first-order kinetic analysis, but with 20 mM purified CpICG in lieu of the 2-MeImpC + ICG combination (Figure 9A; Figure S42, SI). At both pH 6.3 and pH 8.0, the primer +1 product can be clearly observed by 15 min, with its abundance

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792

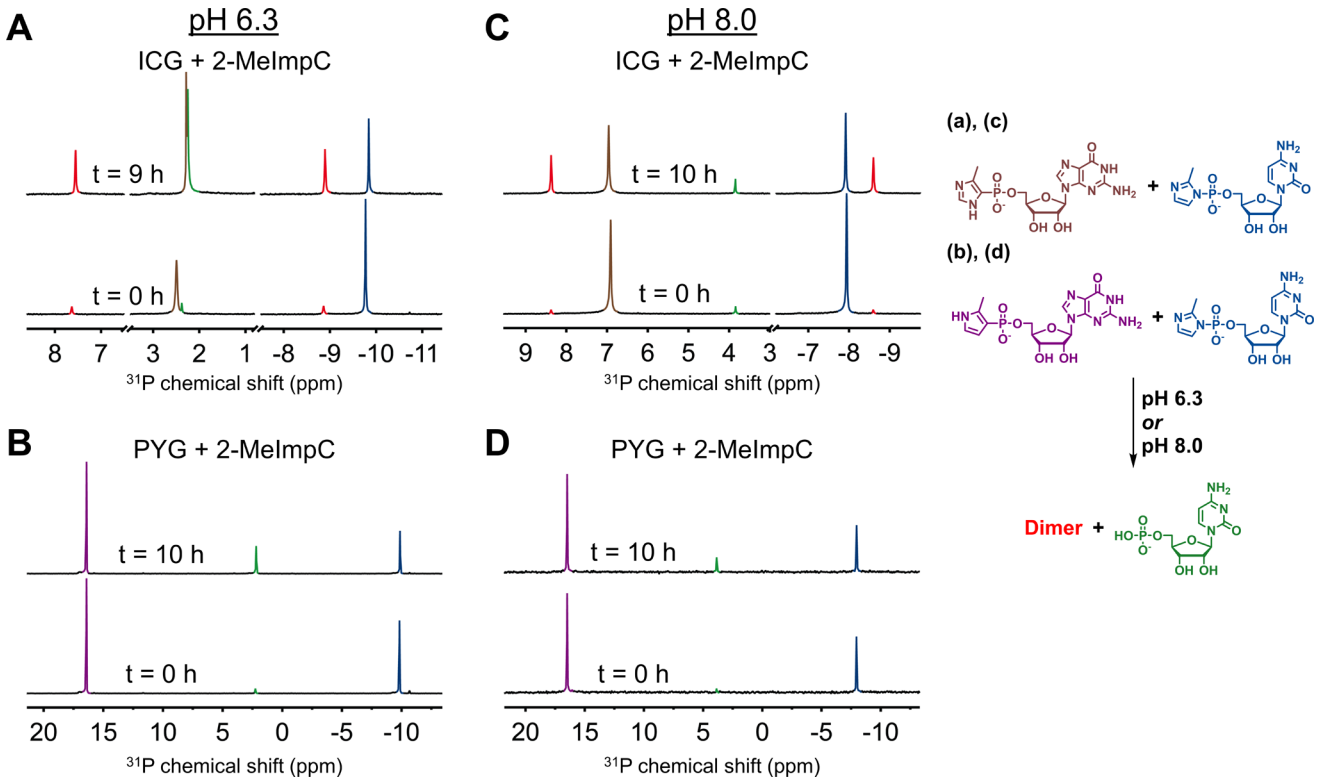


Figure 7. (A) ICG coincubated with 2-MeImpC at pH 6.3. (B) PYG coincubated with 2-MeImpC at pH 6.3. (C) ICG coincubated with 2-MeImpC at pH 8.0. (D) PYG coincubated with 2-MeImpC at pH 8.0. The phosphorus resonances of ICG, 2-MeImpC, CMP, PYG, and GC-dinucleotide were colored as maroon, naval blue, green, purple, and red, respectively. The lack of reactivity between PYG and 2-MeImpC strongly suggests that the ribose hydroxyls and guanine exocyclic amine of ICG do not react significantly with 2-MeImpC. See Figures S21−S24 (SI) for the complete ^{31}P stacked spectrum of the NMR time course analyses depicted in panels (A)−(D), respectively. See Figure S25 (SI) for the complete ^{31}P NMR time course analyses of 2-MeImpC-PZG coincubation. See Figures S37−−S40 (SI) for detailed LCMS time course analyses of the formation and hydrolysis of CpICG at pH 6.3 and 8.0. See Figure S41 (SI) for the fitted rate of CpICG formation and degradation derived from the LCMS time course analyses.

increasing at longer reaction times. This result strongly suggests that CpICG can bind to a cognate RNA template and react with the adjacent primer 3′-hydroxyl group. As a control experiment, we reversed the sequence of the templating nucleotides at the primer +1 and +2 positions (underlined nucleotides, Figure 9A), from 3′-GCCCAA-5′ to 3′-CGAA-5′. If CpICG binds to this "reversed" template, then the unreactive C−P linkage of the imidazole bridge would be pointing toward the primer 3′-OH nucleophile, instead of the labile N−P linkage, and no primer extension should occur. Indeed, reactivity is completely abolished with this template, even at the highest tested CpICG concentration (20 mM, Figure 9A).

To assess the affinity of CpICG for its cognate RNA primer−template duplex, we determined the K_M of this dinucleotide under primer extension conditions (Figure 9B). We measured the pseudo-first-order rate of primer extension as a function of concentration of CpICG at pH 6.3 and 8.0. At pH 6.3, the estimated k_{max} and K_M were 0.28 ± 0.01 h^{-1} and 10.6 ± 0.8 mM, respectively (R^2 = 0.996). At pH 8.0, the estimated k_{max} and K_M were 0.35 ± 0.02 h^{-1} and 12 ± 1 mM, respectively (R^2 = 0.994). The pH change does not significantly alter the affinity of CpICG for its cognate template, and the estimated k_{max} of CpICG-driven primer extension was comparable under both acidic and alkaline conditions, consistent with the similar rates of ICG-catalyzed primer extension with 2-MeImpC at pH 6.3 and 8.0 (k_{obs} = 0.069 ± 0.002 h^{-1} and k_{obs} = 0.066 ± 0.003 h^{-1}, respectively; Figure 4B and Figures S11 and S12, SI).

DISCUSSION

The mechanism by which nonenzymatic RNA polymerization occurs has been of interest for decades. Since the early 90s, it has been known that the presence of an activated monomer downstream of the reacting nucleotide catalyzes primer extension.[26,45] Our laboratory recently confirmed and extended these early results, showing that the presence of an activated 5′-phosphate on the downstream nucleotide (or oligonucleotide) catalyzed the reaction between the primer and the primer-adjacent nucleotide. Noncovalent leaving group−leaving group interactions, or simply steric or Coulombic repulsions between the leaving groups, were thought to catalyze primer extension by bringing the activated phosphate of the reacting monomer closer to the primer 3′-hydroxyl, or helping to orient the leaving group for in-line nucleophilic substitution. Indeed, our previously reported RNA-PZG crystal structures show that the presence of a downstream PZG nucleotide decreases the distance from the primer 3′-OH to the adjacent PZG phosphate from 6.5 to 4.5 Å,[32] and we have obtained similar results in a more recent study with ICG.[46] However, no leaving group-leaving group interactions were observed in either study. If noncovalent leaving group interactions are a critical component of downstream monomer-induced catalysis, PYG, which has a neutral pyrrolyl

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792

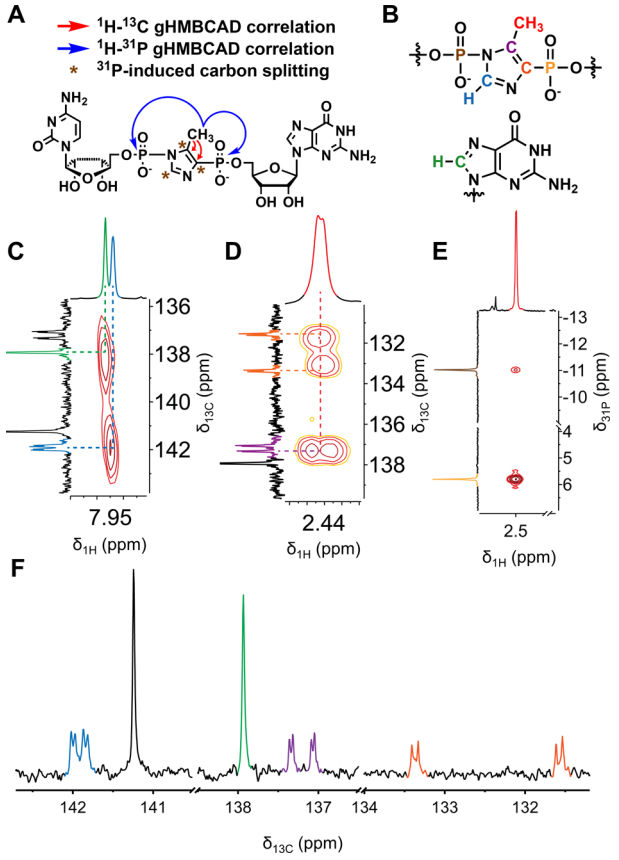


Figure 8. (A) Proposed structure of CpICG, and observed heteronuclear NMR correlations which support the proposed structure. (B) The proton, carbon, and phosphorus atoms associated with the CpICG 5′−5′-imidazole linkage have been assigned to their respective NMR resonances using a combination of 1D and homo- and heteronuclear 2D correlation spectroscopies. The color coding of the atoms is used in panels (C)−(F). (C,D) Expanded view of the ^{1}H−^{13}C coupled (C) gc2HSQCse and (D) gHMBCAD spectrum, covering the spectral region showing the proton−carbon correlations relevant to the 5′−5′-imidazole bridge. Flanked by two phosphorus atoms, the carbon resonances of the 5′−5′-bridge are split as doublet-of-doublets, thus affirming the 5′−5′-linkage regiochemistry of CpICG. (E) Expanded view of the ^{1}H−^{31}P-coupled gHMBCAD spectrum. The exocyclic methyl protons of the bridging 2-methylimidazole are correlated with both imidazole-flanking phosphorus atoms of CpICG via long-range coupling. (F) Expanded spectral view of the CpICG ^{13}C{^{1}H} spectrum that shows the carbon signals of the bridging imidazole. All resonances are clearly split as doublets of doublets, which is consistent with the 5′−5′-imidazole-bridged dinucleotide linkage. See Figures S27−S29 (SI) for the full ^{1}H, ^{13}C, and ^{31}P spectra of CpICG, and Figures S30−S33 (SI) for the full 2D spectrum of ^{1}H−^{1}H gCOSY, ^{1}H−^{13}C gc2HSQCse and gHMBCAD, and ^{1}H−^{31}P gHMBCAD spectra of CpICG.

leaving group mimic (pK_a < 2), should catalyze upstream primer extension to some degree but no catalysis is observed (Figure 4B). Since 4.5 Å is too far for phosphodiester bond formation, other factors must come into play.

Our laboratory has proposed a novel mechanism for primer extension in which the phosphoroimidazolide moieties of two 2-MeImpN first react with each other to form a 5′−5′-imidazolium-bridged dinucleotide intermediate, followed by template binding, primer extension by one nucleotide and regeneration of the downstream 2-MeImpN.[27] The Richert lab has also noted the high reactivity of imidazolium-bridged dinucleotides in primer extension.[23] Primer extension reactions driven by partially purified Cp-MeIm-pC display extremely high (>30 h^{-1}) initial rates of primer extension, at a concentration much lower than the >50 mM concentrations of 2-MeImpC commonly employed,[27] suggesting that catalysis involving a 5′−5′-imidazolium-bridged dinucleotide may be the predominant mechanism of primer extension.

In the present study we synthesized ICG as a close nonhydrolyzable analogue of 2-MeImpG, which we initially hoped to use as a probe of catalytically relevant noncovalent interactions. However, no such interactions have been observed. In addition, we were puzzled by the fact that ICG does act as a modest catalyst of primer extension, but only at mildly acidic pH. We then found that ICG could react with 2-MeImpC to form the corresponding 5′−5′-imidazole-bridged GC dinucleotide (CpICG). Indeed, the rates of ICG-catalyzed primer extensions, carried out with 30 mM of the monomers, can be reproduced with only 3 mM of the purified 5′−5′-imidazole-bridged CpICG (k_{obs} = 0.068 ± 0.004 and 0.074 ± 0.001 h^{-1} for pH 6.3 and 8.0, respectively; Figures S43 and S44, SI), suggesting not only that CpICG is more reactive than the corresponding 2-MeImpC monomer under primer extension conditions, but also that the *in situ* formation of activated 5′−5′-imidazole-bridged dinucleotides could account for the rate of ICG-catalyzed primer extensions shown in Figure 4. These results complement our recent reports[27,33,47] supporting the hypothesis that the formation of 5′−5′-imidazolium-bridged dinucleotides in nonenzymatic RNA primer extensions is an integral step for 2-MeImpN-driven primer extension.

A few characteristics of CpICG are notable. First, the ^{31}P NMR analysis of the products of 2-MeImpC-ICG coincubation suggest that only one regioisomer out of the four theoretically possible GC dinucleotides (imidazole-, 2′−5′- and 3′−5′-phosphodiester-, and nucleobase-linked) is formed. As shown in the 2-MeImpC-PYG coincubation experiment (Figure 7B,D), the ribose hydroxyls and guanine exocyclic amines of the analogues do not detectably react with the electrophilic phosphoroimidazolide moiety of 2-MeImpC. Therefore, formation of the C−P/N−P mixed-linkage dinucleotide occurs with remarkable regioselectivity, with the nonhydrolyzable imidazolylphosphonate of ICG being the exclusive nucleophile during dinucleotide formation. This high regioselectivity simplifies the synthesis and purification of other C−P/N-P mixed-linkage dinucleotide products. Since the identity of the nucleobase is expected to have minimal impact on the yield or regioselectivity of the mixed linkage dinucleotide formation, the preparation of all 16 possible 5′−5′-imidazole-linked, mixed-linkage dinucleotides should be possible. Second, because of the zwitterionic character of the imidazolium bridge, the 2-MeImpN-derived 5′−5′-bridged dinucleotides are inherently reactive, as evidenced by their short half-lives under acidic conditions. Despite careful control of starting material stoichiometry and rigorous purification procedures, preparation and purification of the Cp-MeIm-pC dinucleotide proved to be challenging, affording enriched samples with a maximum of 35% Cp-MeIm-pC.[27] In contrast, the CpICG dinucleotide described herein is quite stable, especially under basic conditions. Even in the presence of a high concentration of Mg^{2+}, CpICG degrades only minimally after 10 h under alkaline conditions (Figures S39 and S40, SI), and lyophilization of an aqueous CpICG solution did not lead to substantial hydrolysis. The stability of CpICG, which enabled our

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792

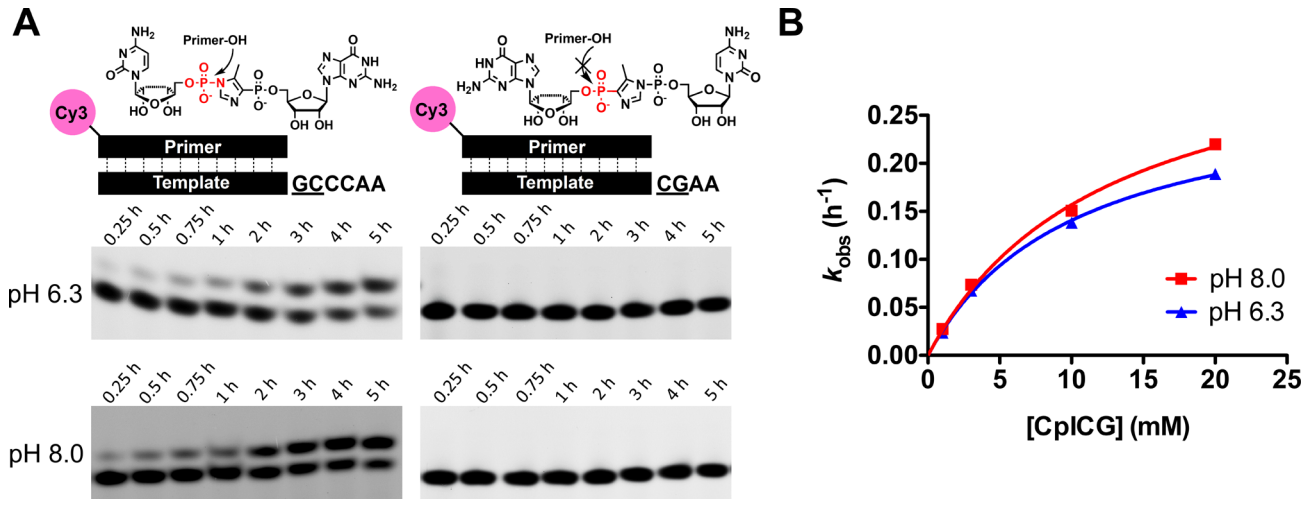


Figure 9. (A) Primer extension using 20 mM CpICG as activated substrate. Primer extension could be observed when the cytidylyl group of CpICG, attached to the bridging imidazole via labile N−P linkage, is adjacent to the primer; no extension could be observed when the guanidylyl group, and the stable C−P linkage, is adjacent to the primer. (B) Saturation curve of the CpICG-driven primer extensions at pH 6.3 and 8.0. See Figures S43 and S44 (SI) for representative denaturing PAGE gel images, first-order rate plots, and the saturation curves of CpICG-driven primer extensions at pH 6.3 and 8.0 at different CpICG concentrations.

extensive 1D and 2D NMR studies of the 5′−5′-imidazole bridge connectivity, will also enable future kinetic, thermodynamic, and crystallographic studies of primer extension. We are currently pursuing crystallographic studies of RNA−CpICG complexes, with the goal of revealing the overall conformation of template-bound CpICG, the noncovalent interactions modulating RNA−CpICG base pairing, the relative orientation and distance between the 5′−5′-imidazole-linkage and the primer 3′-terminus, and other mechanistically relevant structural features.

The most puzzling aspect of ICG catalysis of primer extension with 2-MeImpC was the observation of a moderate catalytic effect at acidic pH, together with the fact that the rate of primer extension was slow but almost pH independent from 6.3 to 8.0. Here we present a possible explanation for these observations. CpICG-driven primer extension requires at least two major mechanistic steps: (1) nucleophilic attack of the primer 3′-OH on the imidazole bridge from the labile N−P bond side, and (2) displacement of the imidazolylphosphonate group to liberate a neutral ICG analogue. At high pH, the 5′−5′-bridging 2-methylimidazole of CpICG is uncharged; direct nucleophilic displacement of the imidazole phosphonate bridge, without prior imidazole protonation or concomitant proton transfer, would result in an unfavorable anionic imidazolyl-phosphonate leaving group. Therefore, the imidazolyl group must first be protonated and become cationic, before nucleophilic attack by the primer 3′-OH takes place. Under acidic conditions, the 5′−5′-imidazole bridge of CpICG is more likely to be protonated (estimated pK_a ≈5.1, Figure S45, SI), but deprotonation of the primer 3′-OH will be correspondingly more difficult. We suggest that the need to deprotonate the primer 3′-OH while simultaneously protonating the imidazolyl-phosphonate leads to a low primer extension rate regardless of overall pH, together with a catalytic effect of ICG at acidic pH when the intermediate can be protonated. On the other hand, preliminary data suggested that PZG does react with 2-MeImpC to give the 5′−5′-pyrazole-bridged CpPZG (Figure S25, SI), yet PZG is unable to catalyze primer extensions to any significant extent (Figure 4B). This lack of reactivity may be due to the low protonation pK_a of the 5′−5′-pyrazole bridge (likely to be 1.94 or less; Figure 1 and Figure S2, SI), which renders cationic activation of the pyrazole bridge unlikely when primer extension is carried out at pH ≥5.5.

In conclusion, we have described the synthesis and properties of ICG, the closest nonhydrolyzable 2-MeImpG analogue reported to date. ICG acts as a modest catalyst of primer extension under acidic conditions, primarily via the formation of a 5′−5′-imidazole-bridged dinucleotide which carries both N−P and C−P linkages. This relatively stable mixed-linkage dinucleotide serves as a substrate in primer extension reactions, consistent with our recent mechanistic proposal that the formation of 5′−5′-imidazolium-bridged dinucleotides is an integral step for 2-MeImpN-driven primer extensions. The affinity of these dinucleotides for cognate and noncognate RNA primer−templates, as well as the crystal structures of RNA−dinucleotide complexes, are currently being determined.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b11623.

Figures S1−S45, Table S1, methods, and references (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*szostak@molbio.mgh.harvard.edu

ORCID

Chun Pong Tam: 0000-0001-6381-9011
Lijun Zhou: 0000-0002-0393-4787
Albert C. Fahrenbach: 0000-0002-8315-8836
Travis Walton: 0000-0001-6812-1579
Jack W. Szostak: 0000-0003-4131-1203

Author Contributions

[∇]C.P.T. and L.Z. contributed equally.

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge Mr. Gregory J. Heffron and Dr. Charles A. Sheehan, of Harvard Medical School East Quadrangle Nuclear Magnetic Resonance Core Facility, for their assistance with NMR spectroscopy. We thank Dr. Victor S. Lelyveld for assistance with high-resolution liquid chromatography-coupled mass spectrometry. We thank Prof. Seung Soo Oh, Dr. Daniel Duzdevich, Dr. Kyle Strom, Dr. Li Li, and Dr. Derek O'Flaherty for insightful discussions and critiques of the manuscript. J.W.S. is an Investigator of the Howard Hughes Medical Institute. This work was supported in part by a grant from the NSF (CHE-1607034) to J.W.S. and a grant (290363) from the Simons Foundation to J.W.S.

■ REFERENCES

(1) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. *Nature* **2001**, *409*, 387.
(2) Blain, J. C.; Szostak, J. W. *Annu. Rev. Biochem.* **2014**, *83*, 615.
(3) Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99.
(4) Robertson, M. P.; Joyce, G. F. *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003608.
(5) Szostak, J. W. *J. Syst. Chem.* **2012**, *3*, 2.
(6) Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. *eLife* **2016**, *5*, e17756.
(7) Kanavarioti, A.; Hurley, T. B.; Baird, E. E. *J. Mol. Evol.* **1995**, *41*, 161.
(8) Kanavarioti, A.; Bernasconi, C. F.; Baird, E. E. *J. Am. Chem. Soc.* **1998**, *120*, 8575.
(9) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3340.
(10) Prabahar, K. J.; Ferris, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 4330.
(11) Vogel, S. R.; Deck, C.; Richert, C. *Chem. Commun.* **2005**, *39*, 4922.
(12) Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E. *Science* **1968**, *161*, 387.
(13) Inoue, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1981**, *103*, 7666.
(14) Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. *J. Am. Chem. Soc.* **2017**, *139*, 1810.
(15) Deck, C.; Jauker, M.; Richert, C. *Nat. Chem.* **2011**, *3*, 603.
(16) Heuberger, B. D.; Pal, A.; Del Frate, F.; Topkar, V. V.; Szostak, J. W. *J. Am. Chem. Soc.* **2015**, *137*, 2769.
(17) Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A. *J. Am. Chem. Soc.* **2010**, *132*, 5880.
(18) Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1982**, *162*, 201.
(19) Bridson, P. K.; Orgel, L. E. *J. Mol. Biol.* **1980**, *144*, 567.
(20) Giurgiu, C.; Li, L.; O'Flaherty, D. K.; Tam, C. P.; Szostak, J. W. *J. Am. Chem. Soc.* **2017**, *139*, 16741.
(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) Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2015**, *137*, 6373.
(23) Kervio, E.; Sosson, M.; Richert, C. *Nucleic Acids Res.* **2016**, *44*, 5504.
(24) Oró, J.; Basile, B.; Cortes, S.; Shen, C.; Yamrom, T. *Origins Life* **1984**, *14*, 237.
(25) Fahrenbach, A. C.; Giurgiu, C.; Tam, C. P.; Li, L.; Hongo, Y.; Aono, M.; Szostak, J. W. *J. Am. Chem. Soc.* **2017**, *139*, 8780.
(26) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 5496.
(27) Walton, T.; Szostak, J. W. *J. Am. Chem. Soc.* **2016**, *138*, 11996.
(28) Korhonen, H. J.; Conway, L. P.; Hodgson, D. R. W. *Curr. Opin. Chem. Biol.* **2014**, *21*, 63.
(29) Temiakov, D.; Patlan, V.; Anikin, M.; McAllister, W. T.; Yokoyama, S.; Vassylyev, D. G. *Cell* **2004**, *116*, 381.
(30) Upton, T. G.; Kashemirov, B. A.; McKenna, C. E.; Goodman, M. F.; Prakash, G. K. S.; Kultyshev, R.; Batra, V. K.; Shock, D. D.; Pedersen, L. C.; Beard, W. A.; Wilson, S. H. *Org. Lett.* **2009**, *11*, 1883.
(31) Desai, K. K.; Raines, R. T. *Biochemistry* **2012**, *51*, 1333.
(32) Zhang, W.; Tam, C. P.; Wang, J.; Szostak, J. W. *ACS Cent. Sci.* **2016**, *2*, 916.
(33) Zhang, W.; Tam, C. P.; Walton, T.; Fahrenbach, A. C.; Birrane, G.; Szostak, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7659.
(34) Van Den Berge, E.; Robiette, R. *J. Org. Chem.* **2013**, *78*, 12220.
(35) Kalek, M.; Ziadi, A.; Stawinski, J. *Org. Lett.* **2008**, *10*, 4637.
(36) McKenna, C. E.; Higa, M. T.; Cheung, N. H.; McKenna, M.-C. *Tetrahedron Lett.* **1977**, *18*, 155.
(37) Taylor, S. D.; Mirzaei, F.; Bearne, S. L. *J. Org. Chem.* **2008**, *73*, 1403.
(38) Rabenstein, D. L.; Sayer, T. L. *Anal. Chem.* **1976**, *48*, 1141.
(39) Bezençon, J.; Wittwer, M. B.; Cutting, B.; Smieško, M.; Wagner, B.; Kansy, M.; Ernst, B. *J. Pharm. Biomed. Anal.* **2014**, *93*, 147.
(40) Krezel, A.; Bal, W. *J. Inorg. Biochem.* **2004**, *98*, 161.
(41) Kanavarioti, A.; Rosenbach, M. T. *J. Org. Chem.* **1991**, *56*, 1513.
(42) Perrin, D. D. *Dissociation constants of organic bases in aqueous solution*; Butterworths: London, 1965.
(43) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N. *Anal. Biochem.* **1989**, *179*, 131.
(44) Schmidtchen, F. P. In *Analytical Methods in Supramolecular Chemistry*; Schalley, C., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2007; pp 55−78.
(45) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 7963.
(46) Zhang, W.; Tam, C. P.; Zhou, L.; Oh, S. S.; Wang, J.; Szostak, J. W. Manuscript submitted.
(47) Walton, T.; Szostak, J. W. *Biochemistry* **2017**, *56*, 5739.

DOI: 10.1021/jacs.7b11623
J. Am. Chem. Soc. 2018, 140, 783−792