Supporting Information

A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension

Travis Walton and Jack W. Szostak*

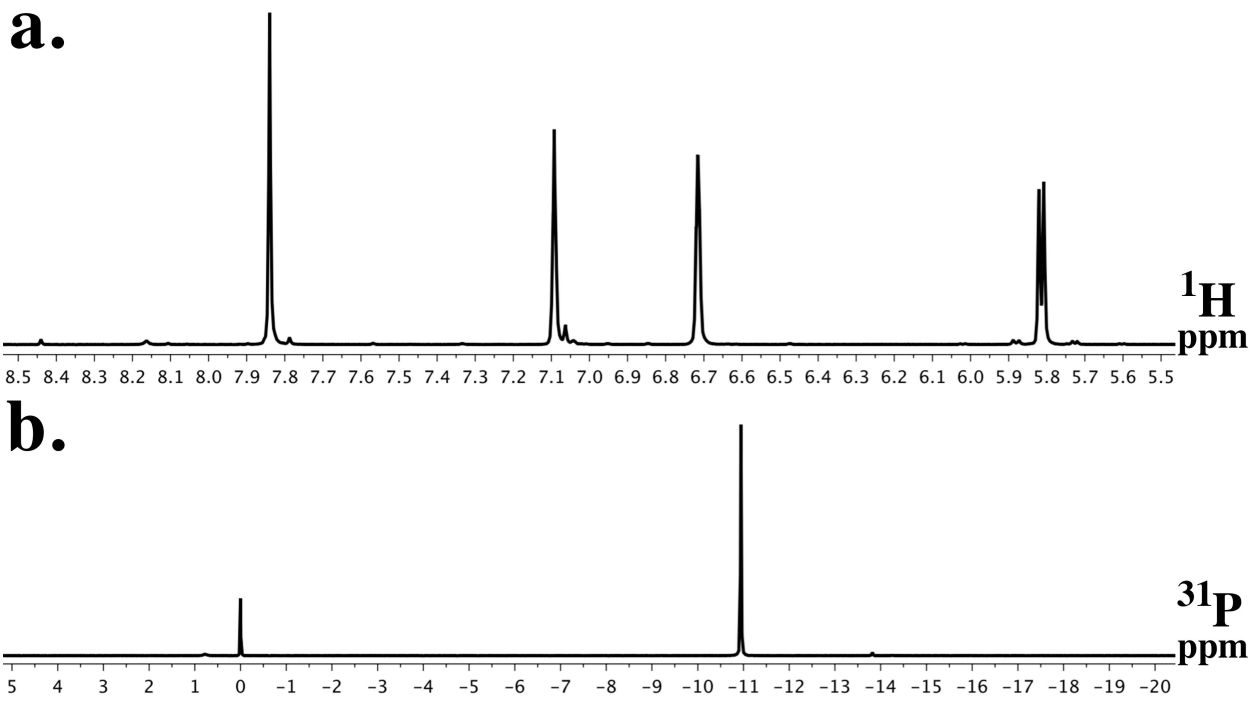

Figure S1. The imidazolium bridged di-guanosine intermediate **2a** is barely detectable by ^{1}H and ^{31}P NMR at basic pH. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2-MeImpG **1a** incubated 57 minutes at pD 8.49. For 2-MeImpG, the peak at 7.84 ppm corresponds to H8, the peaks at 7.09 and 6.72 ppm correspond to the aromatic 2-methylimidazole protons, and the doublet at 5.82 ppm corresponds to H1′. The small peaks at 8.16 (H8) and 5.88 (H1′) correspond to GMP. The small peaks at 7.79 and 5.73 ppm are attributed to the imidazolium-bridged intermediate. By integration, the concentration of this intermediate is < 500 μM. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2-MeImpG incubated 94 minutes at pD 8.49. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -10.94 ppm corresponds to 2-MeImpG. The small peak at 0.78 ppm corresponds to GMP. The small peak at -13.82 ppm is attributed to an intermediate, due to the identical location of this peak at neutral pD (Fig. S3). Peak assignments for GMP were confirmed by addition of standards.

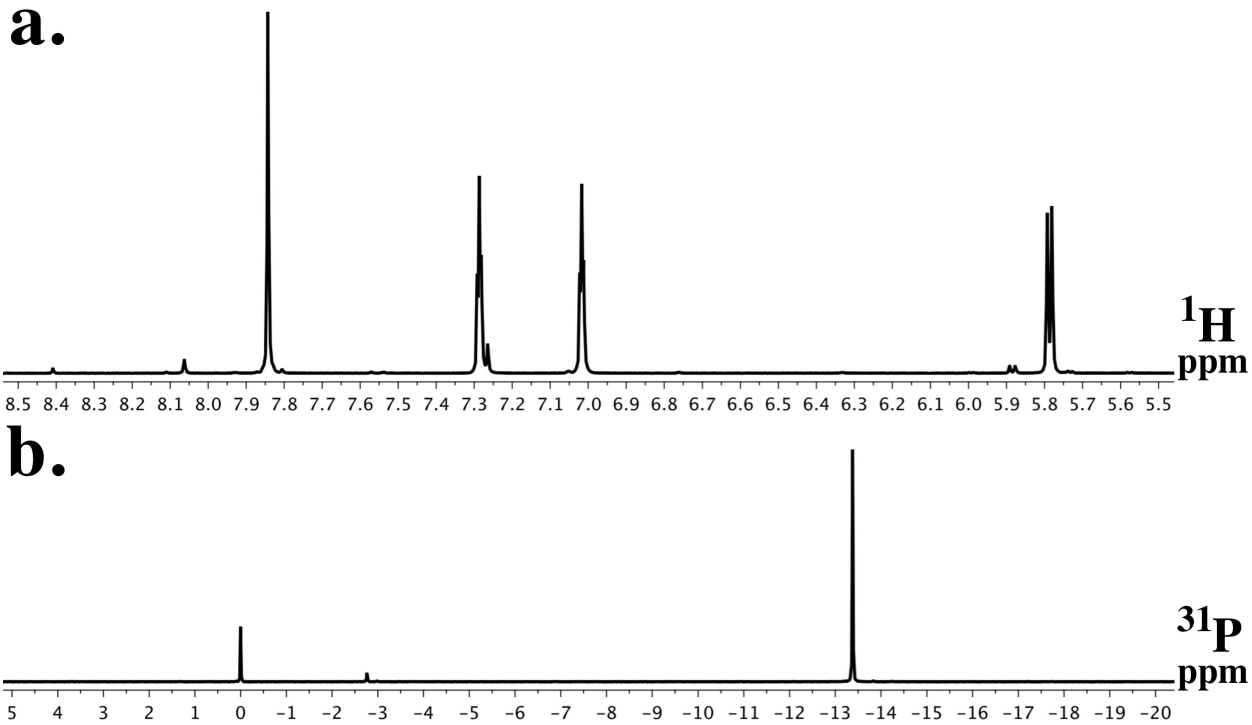

Figure S2. The imidazolium bridged di-guanosine intermediate **2a** is barely detectable by ^{1}H and ^{31}P NMR at acidic pH. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2-MeImpG **1a** incubated 55 minutes at pD 4.62. The peaks at 7.29 and 7.02 ppm correspond to the 2-methylimidazole aromatic protons of 2-MeImpG, which have shifted compared to pD 8.49 (Fig. S1) due to the change in the ionization state of the 2-methylimidazole. The small peak at 7.26 ppm corresponds to the aromatic protons of free 2-methylimdazole. By integration, the concentration of **2a** is < 500 μM. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2-MeImpG incubated 94 minutes at pD 4.62. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -13.38 ppm corresponds to 2-MeImpG. Peak assignments for GMP were confirmed by addition of standards.

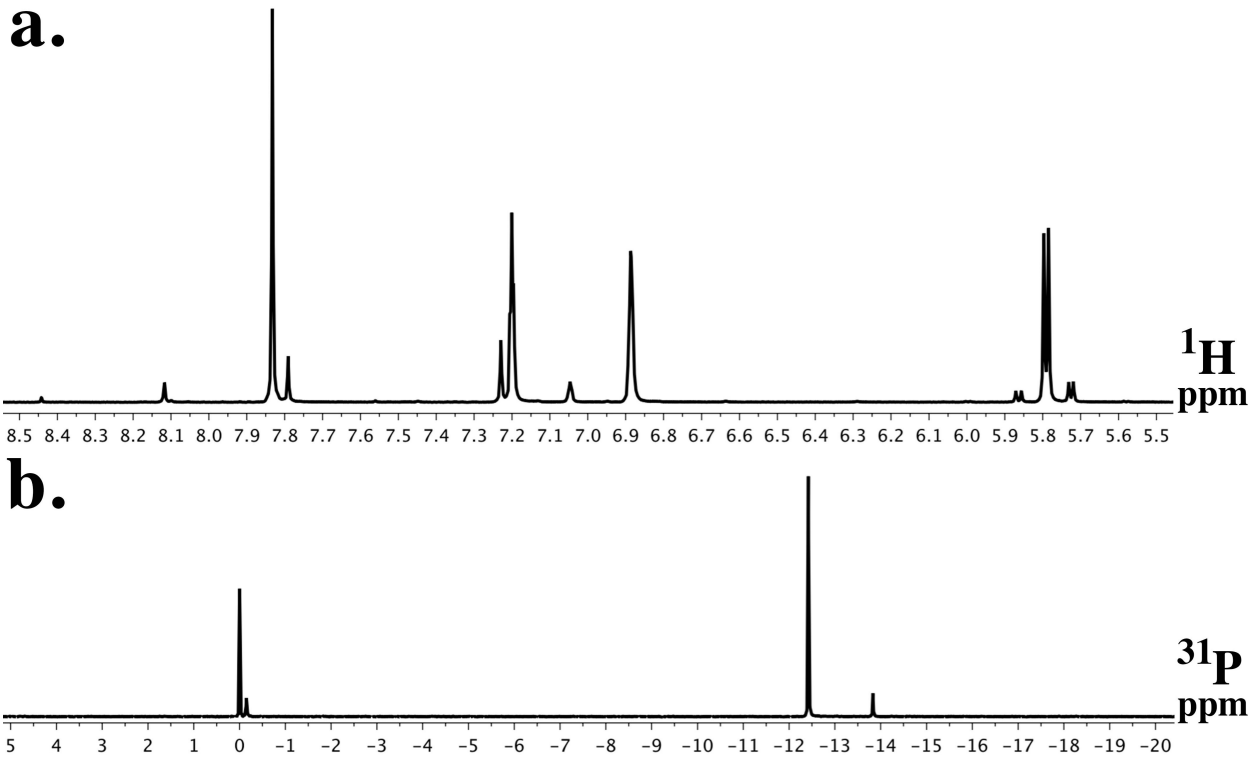

Figure S3. The imidazolium bridged di-guanosine intermediate **2a** is clearly observed at neutral pH. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2-MeImpG **1a** incubated 55 minutes at pD 7.06. The peaks at 7.20 and 6.89 ppm correspond to the aromatic 2-methylimidazole protons of 2-MeImpG, which have shifted compared to pD 8.49 and pD 4.62 spectra (Fig. S1–2) due to the altered protonation state of the 2-methylimidazole. On the basis of a ^{1}H spectrum of 2-methylimidazole taken at pD 7.02, the peak at 7.23 ppm corresponds to the aromatic protons of free 2-methylimidazole. Due to proximity, we attribute the peak at 7.79 ppm to the H8 of **2a**, the peak at 7.05 ppm to the aromatic 2-methylimidazole protons of **2a**, and the peak at 5.73 ppm to the H1′ of **2a**. By integration, the concentration of intermediate **2a** is 2.5 mM. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2-MeImpG incubated 117 minutes at pD 7.06. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -12.42 ppm corresponds to 2-MeImpG. The peak at -0.15 ppm corresponds to GMP. The peak at -13.83 ppm is attributed to intermediate **2a**, on the basis of our kinetic analysis of the primer extension reaction (Fig. 2–3).

In addition, the peak at -13.83 ppm does not shift with pH, as expected since the bridging imidazolium cannot change ionization state with pH.

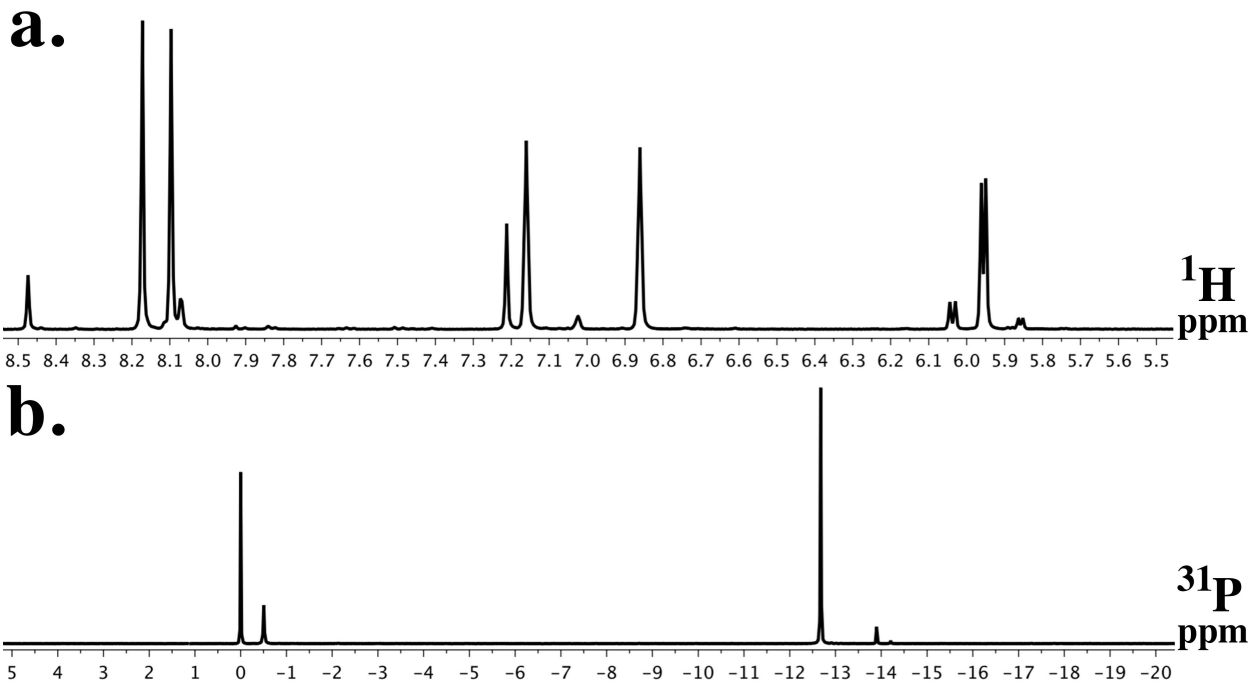

Figure S4. Formation of an imidazolium bridged di-adenosine intermediate is observed at neutral pH for 2-MeImpA. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2-MeImpA incubated 60 minutes at pD 6.93. Based upon our analysis of 2-MeImpG (Fig. S3), the peak at 7.02 ppm corresponds to the aromatic 2-methylimidazolium protons of the di-adenosine intermediate. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2-MeImpA incubated 99 minutes at pD 6.93. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -0.50 ppm corresponds to adenosine-5′-monophosphate. The peak at -12.68 ppm corresponds to 2-MeImpA. The peak at -13.90 ppm corresponds to the imidazolium-bridged di-adenosine intermediate, based upon proximity to the peak for intermediate **2a** (Fig. S3).

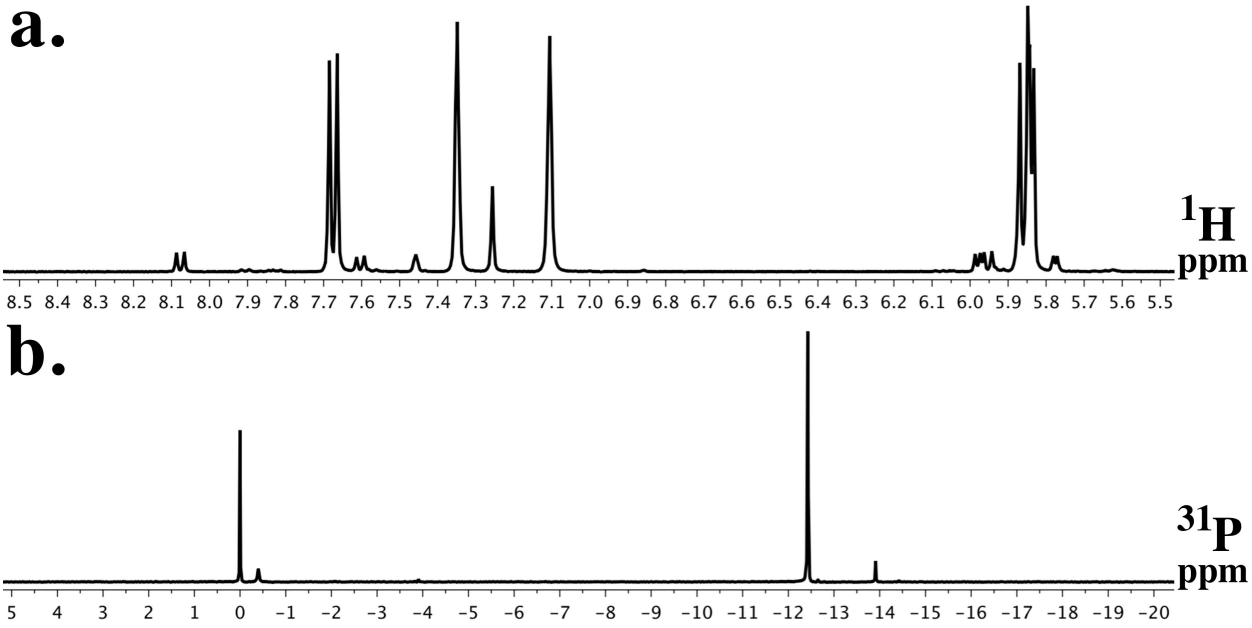

Figure S5. Formation of an imidazolium bridged di-uridine intermediate is observed at neutral pH for 2-MeImpU. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2-MeImpU incubated 57 minutes at pD 6.92. Relative to the purine monomers (Fig. S3–4), the peaks for the aromatic 2-methylimidazole protons of 2-MeImpU are shifted downfield to 7.35 and 7.11 ppm. For all ^{1}H spectra, the peak at 7.26 ppm corresponds to the aromatic protons of free 2-methylimidazole. The peak at 7.46 ppm corresponds to the aromatic 2-methylimidazolium protons of an imidazolium-bridged di-uridine intermediate, on the basis of proximity to other H_{imid} peaks and the reduced intensity of this peak at pD 8.80. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2-MeImpU incubated 91 minutes at pD 6.92. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -0.40 ppm corresponds to uridine-5′-monophosphate. The peak at -12.43 ppm corresponds to 2-MeImpU. The peak at -13.91 ppm corresponds to the di-uridine intermediate, on the basis of proximity to the peak of intermediate **2a** (Fig. S3).

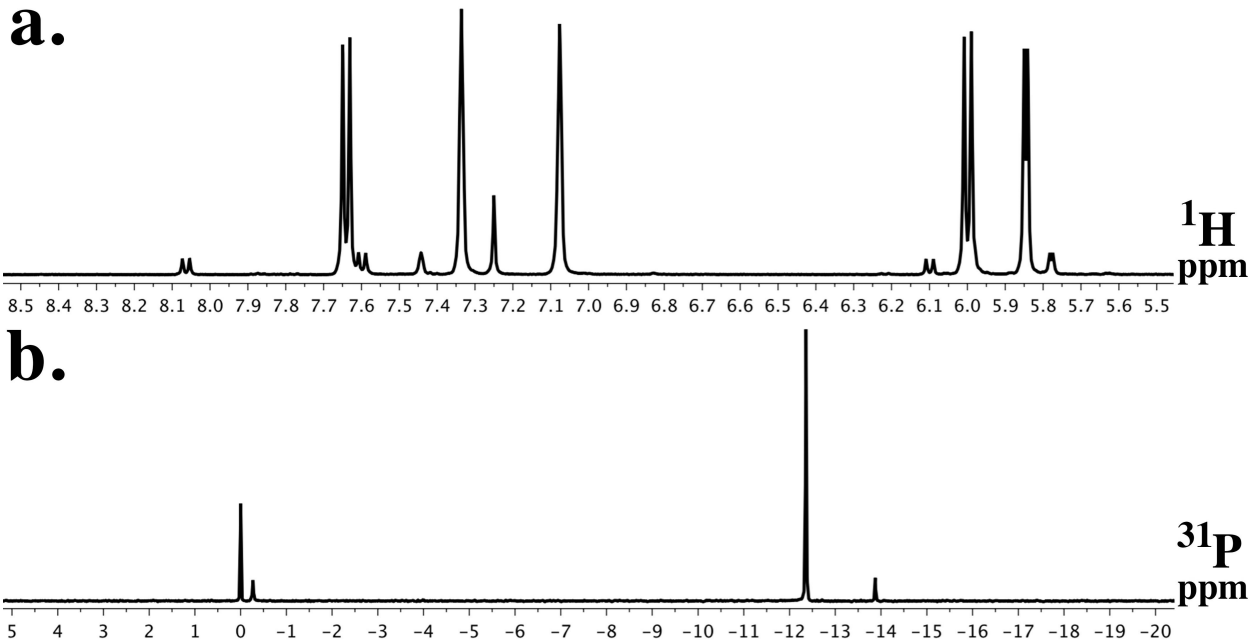

Figure S6. Formation of an imidazolium bridged di-cytidine intermediate is observed at neutral pH for 2-MeImpC. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2-MeImpC monomer incubated 54 minutes at pD 7.02. Similar to the spectrum of 2-MeImpU (Fig. S5), we observed a peak at 7.46 ppm that corresponds to the aromatic 2-methylimidazole protons of the proposed imidazolium-bridged intermediate. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2-MeImpC incubated 89 minutes at pD 7.02. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -0.27 ppm corresponds to cytidine-5′-monophosphate. The peak at -12.36 ppm corresponds to 2-MeImpC. The peak at -13.88 ppm corresponds to the imidazolium-bridged intermediate, on the basis of our analysis of material enriched for this molecule (Fig. 4).

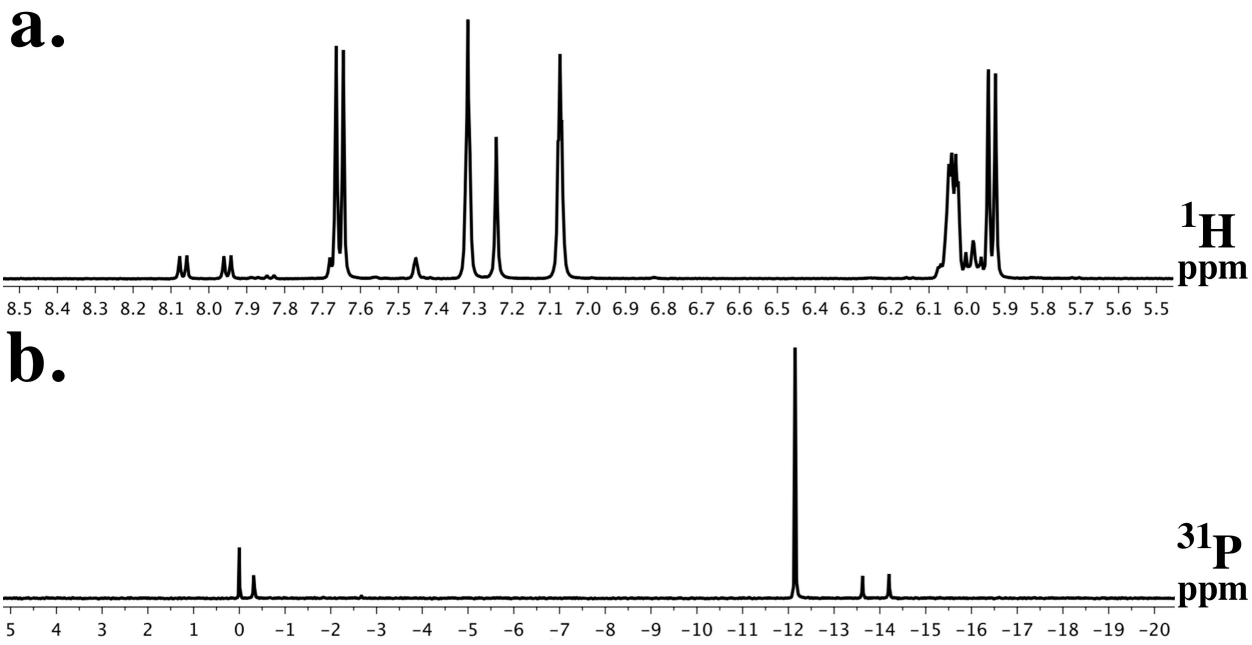

Figure S7. Formation of an imidazolium bridged dinucleotide is observed at neutral pH for 2′,3′-dideoxy-2-MeImpC. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2′,3′-dideoxy-2-MeImpC monomer incubated 60 minutes at pD 7.03. Similar to 2-MeImpU and 2-MeImpC (Fig. S5–6), the peak at 7.45 ppm is attributed to the aromatic protons of the 2-methylimidazole group on an imidazolium-bridged intermediate. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2′,3′-dideoxy-2-MeImpC incubated 95 minutes at pD 7.02. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -0.32 ppm corresponds to 2′,3′-dideoxy-cytidine-5′-monophosphate. The peak at -12.15 ppm corresponds to 2′,3′-dideoxy-2-MeImpC. The peak at -13.62 ppm corresponds to an imidazolium-bridged intermediate, on the basis of proximity of this peak to other imidazolium-bridged dinucleotides (Fig. S6). The peak at -14.20 corresponds to di-(2′,3′-dideoxy-cytidine)-5′,5′-pyrophosphate, on the basis of proximity to the phosphorus resonance for CppC (Fig. S9).

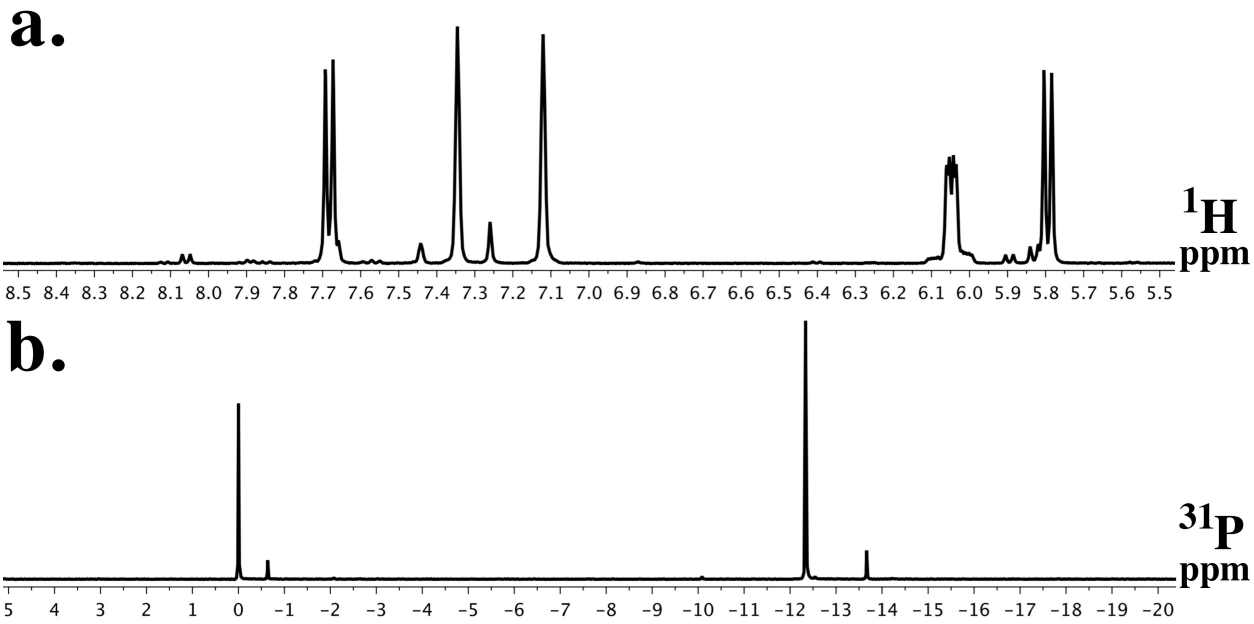

Figure S8. Formation of an imidazolium bridged intermediate is observed at neutral pH for 2′,3′-dideoxy-2-MeImpU. (a) ^{1}H (400 MHz) NMR spectrum of 50 mM 2′,3′-dideoxy-2-MeImpU monomer incubated 55 minutes at pD 6.98. Similar to 2-MeImpU and 2-MeImpC (Fig. S5–6), the peak at 7.44 ppm is attributed to the aromatic protons of the 2-methylimidazole group on the imidazolium-bridged intermediate. (b) ^{31}P (161 MHz) NMR spectrum of 50 mM 2′,3′-dideoxy-2-MeImpU incubated 82 minutes at pD 6.98. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -0.64 ppm corresponds to 2′,3′-dideoxy-uridine-5′-monophosphate. The peak at -12.34 ppm corresponds to 2′,3′-dideoxy-2-MeImpU. The peak at -13.67 ppm corresponds to the imidazolium-bridged intermediate, on the basis of proximity to the corresponding resonance for the uridine ribonucleotide imidazolium-bridged dinucleotide (Fig. S5).

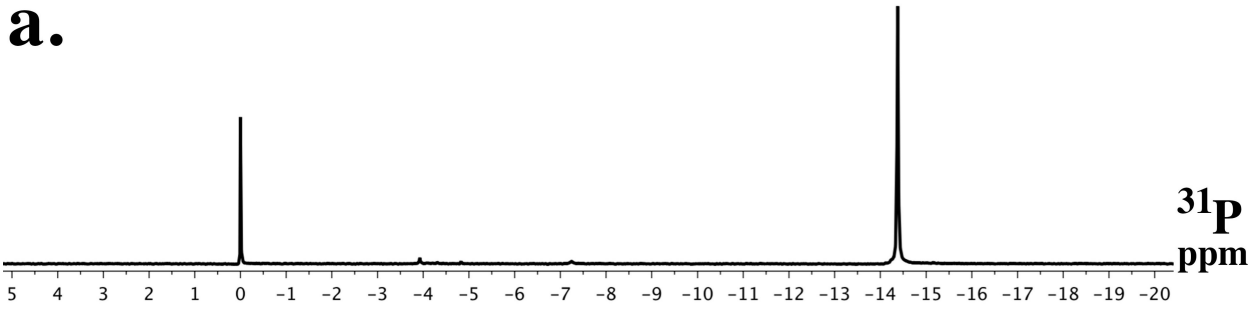

Figure S9. CppC has a ^{31}P chemical shift distinct from that of the imidazolium-bridged intermediate. (a) ^{31}P (161 MHz) NMR spectrum of 26 mM CppC at pD 6.95. The spectrum is referenced to trimethyl phosphate at 0.00 ppm. The peak at -14.38 ppm corresponds to CppC.

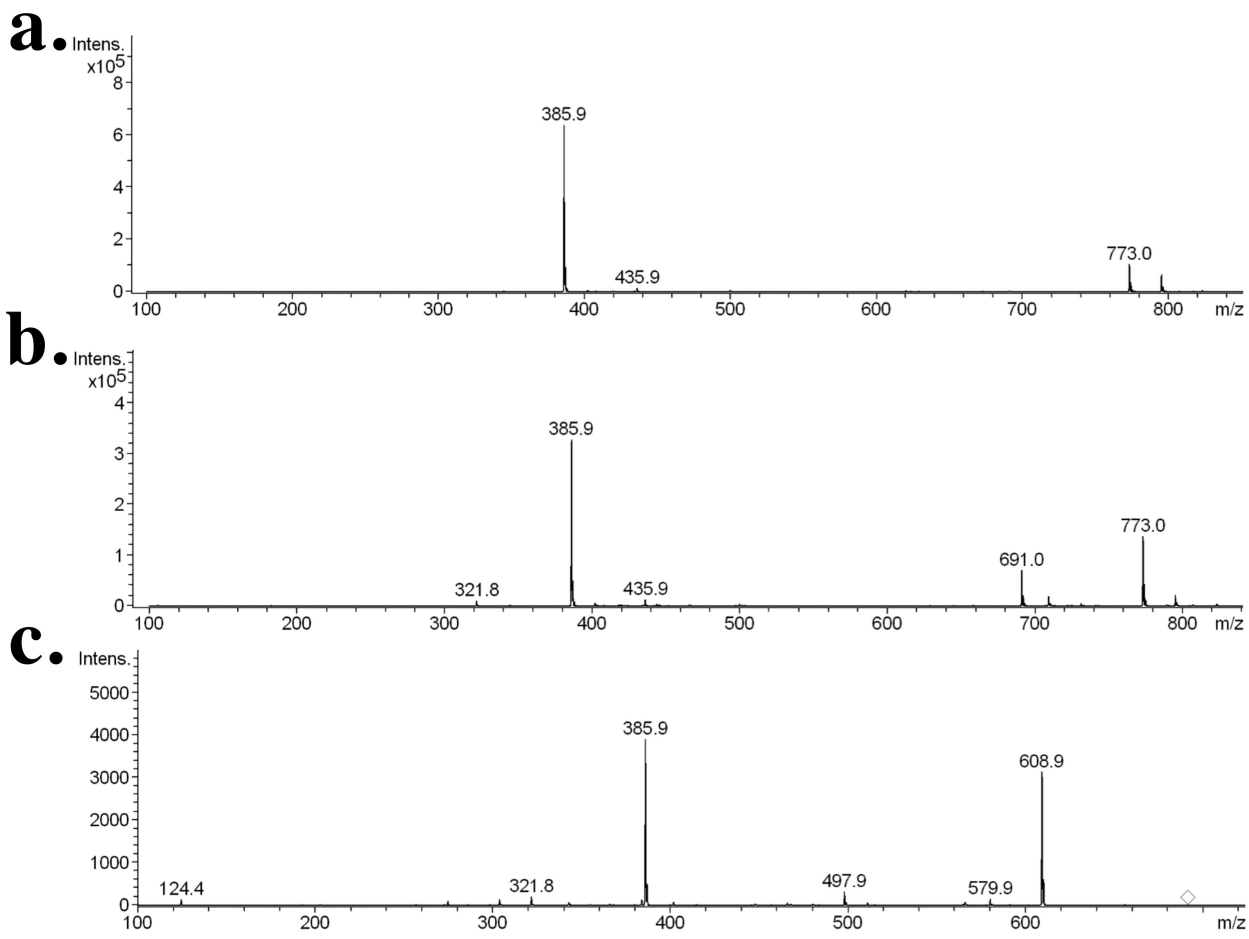

Figure S10. Negative-ion mode ESI-MS analysis of a covalent intermediate formed by incubation of 2-MeImpC at pH 7. (a) ESI-MS spectrum of 100 mM 2-MeImpC incubated at pH 11.3 for 30 minutes. (b) ESI-MS spectrum of 100 mM 2-MeImpC incubated at pH 7.0 for 30 minutes. The new peak at m/z = 691.0 corresponds to the mass of the imidazolium-bridged dinucleotide. (c) MS^2 fragmentation spectrum of m/z = 691, the mass of the proposed intermediate. m/z = 321.8 corresponds to cytidine-5′-monophosphate, m/z = 385.9 to 2-MeImpC, m/z = 497.9 to loss of 2-methylimidazole and cytosine from m/z = 691, m/z = 579.9 to loss of cytosine from m/z = 691, m/z = 608.9 to loss of 2-methylimidazole from m/z = 691, m/z = 691.0 to a dinucleotide covalently bonded to 2-methylimidazole, and m/z = 773.0 to a noncovalent $(2M + 1H^+)^-$ dimer of 2-MeImpC.

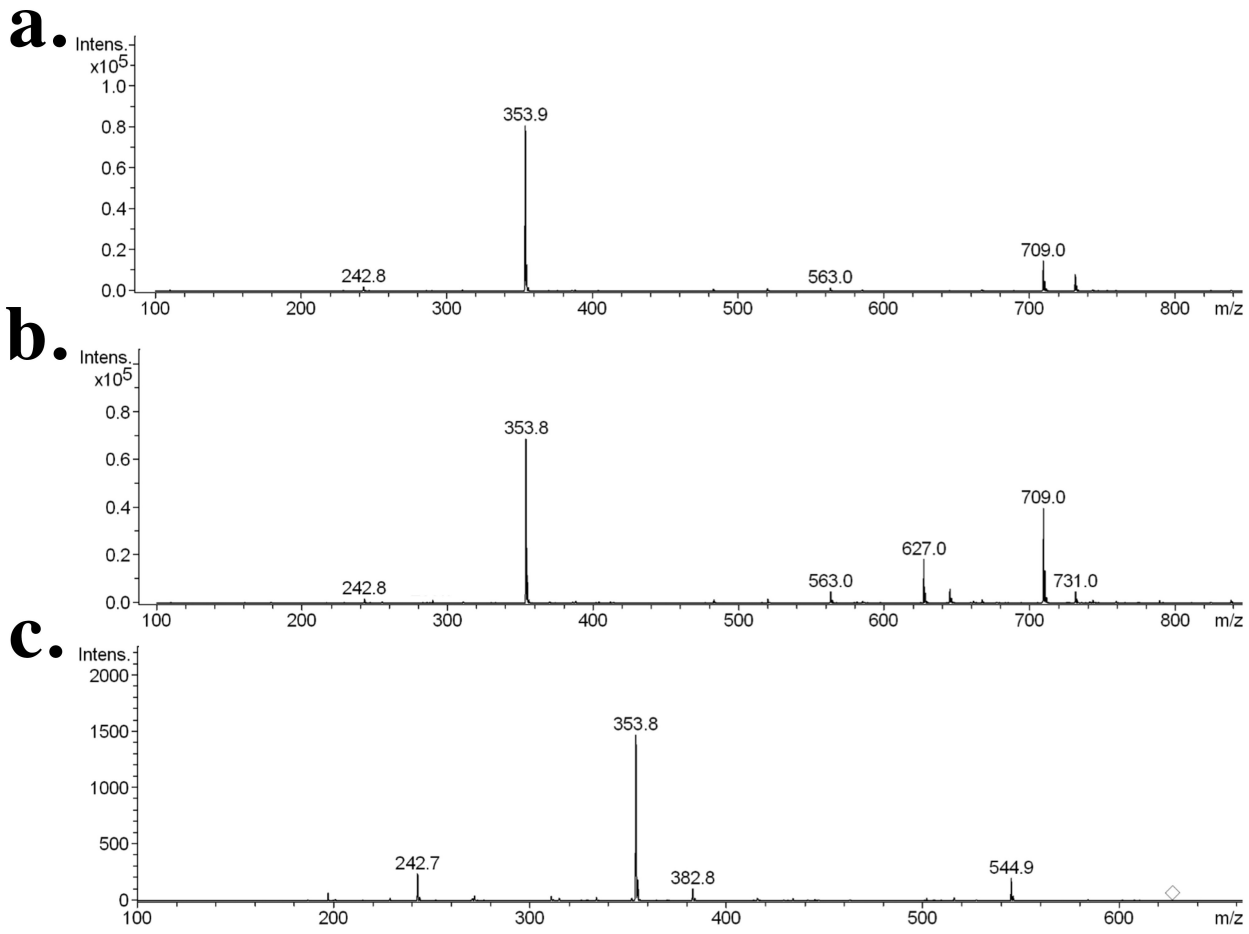

Figure S11. Negative-ion mode ESI-MS analysis of a covalent intermediate formed by incubation of 2′,3′-dideoxy-2-MeImpC at pH 7. (a) ESI-MS spectrum of 100 mM 2′,3′-dideoxy-2-MeImpC incubated at pH 9.2 for 30 minutes. (b) ESI-MS spectrum of 100 mM 2′,3′-dideoxy-2-MeImpC incubated at pH 7.0 for 30 minutes. The new peak at m/z = 627.0 corresponds to the mass of an imidazolium-bridged dinucleotide. (c) MS^2 fragmentation spectrum of m/z = 627, the mass of the proposed intermediate. m/z = 242.8 corresponds to loss of cytosine from m/z = 353.8, m/z = 353.8 to 2′,3′-dideoxy-2-MeImpC, m/z = 382.8 to 2′,3′-dideoxy-CMP covalently linked to cytosine, m/z = 563.0 to CppC, m/z = 627.0 to a covalent dimer of 2′,3′-dideoxy-CMP bonded to 2-methylimidazole, and 709.0 to a noncovalent $(2M + 1H^+)^-$ dimer of 2′,3′-dideoxy-2-MeImpC.

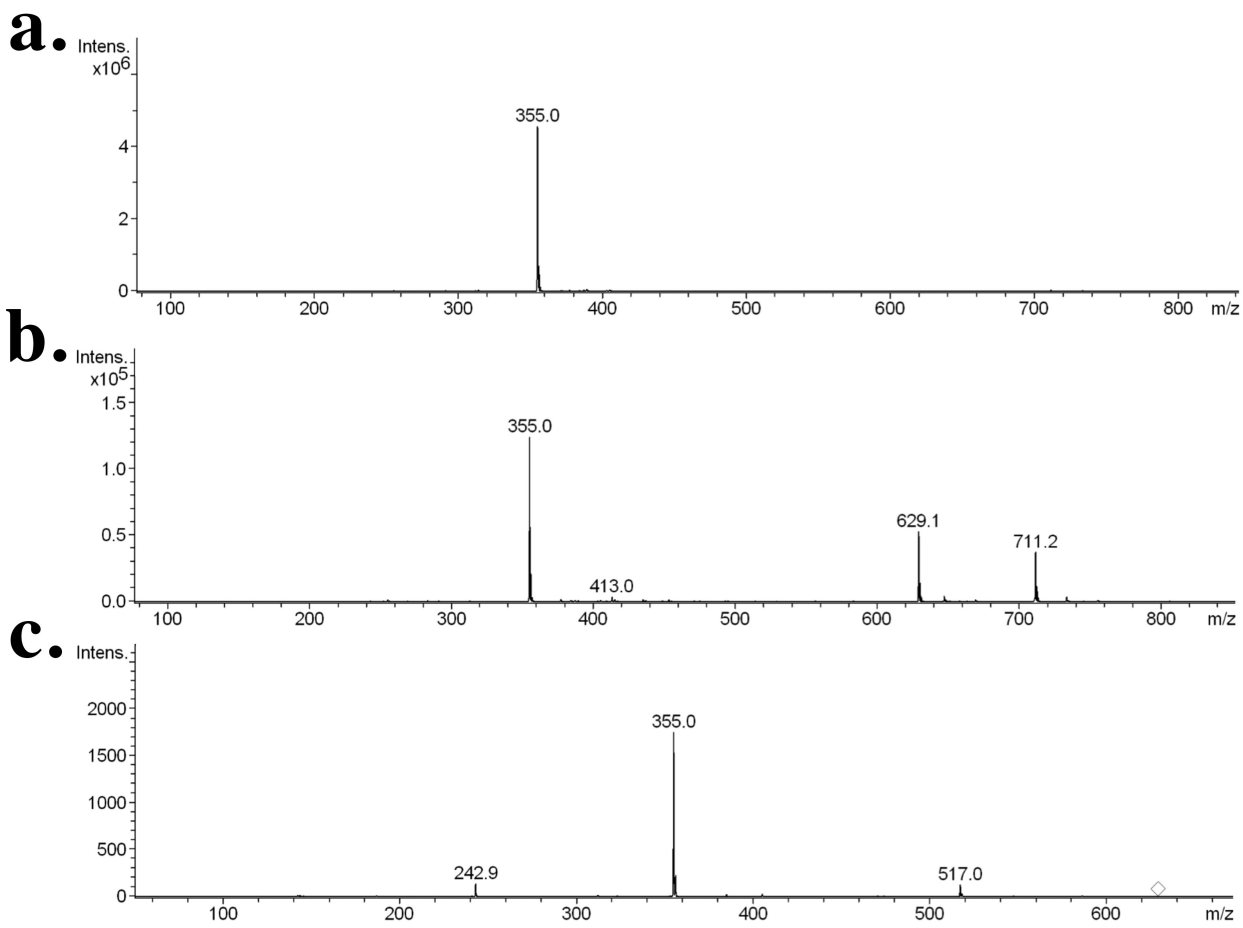

Figure S12. Negative-ion mode ESI-MS analysis of a covalent intermediate formed by incubation of 2′,3′-dideoxy-2-MeImpU at pH 7. (a) ESI-MS spectrum of 100 mM 2′,3′-dideoxy-2-MeImpU incubated at pH 8.6 for 30 minutes. (b) ESI-MS spectrum of 100 mM 2′,3′-dideoxy-2-MeImpU incubated at pH 7.0 for 30 minutes. The new peak at m/z = 629.1 corresponds to the mass of an imidazolium-bridged dinucleotide. (c) MS^2 fragmentation spectrum of m/z = 629, the mass of the proposed intermediate. m/z = 242.9 corresponds to loss of uridine from 2′,3′-dideoxy-2-MeImpU, m/z = 355.0 to 2′,3′-dideoxy-2-MeImpU, m/z = 517.0 to loss of uridine from m/z = 629.1, m/z = 629.1 to a covalent dimer of 2′,3′-dideoxy-UMP covalently linked to 2-methylimidazole, and 711.2 to a noncovalent $(2M + 1H^+)^-$ dimer of 2′,3′-dideoxy -2-MeImpU.

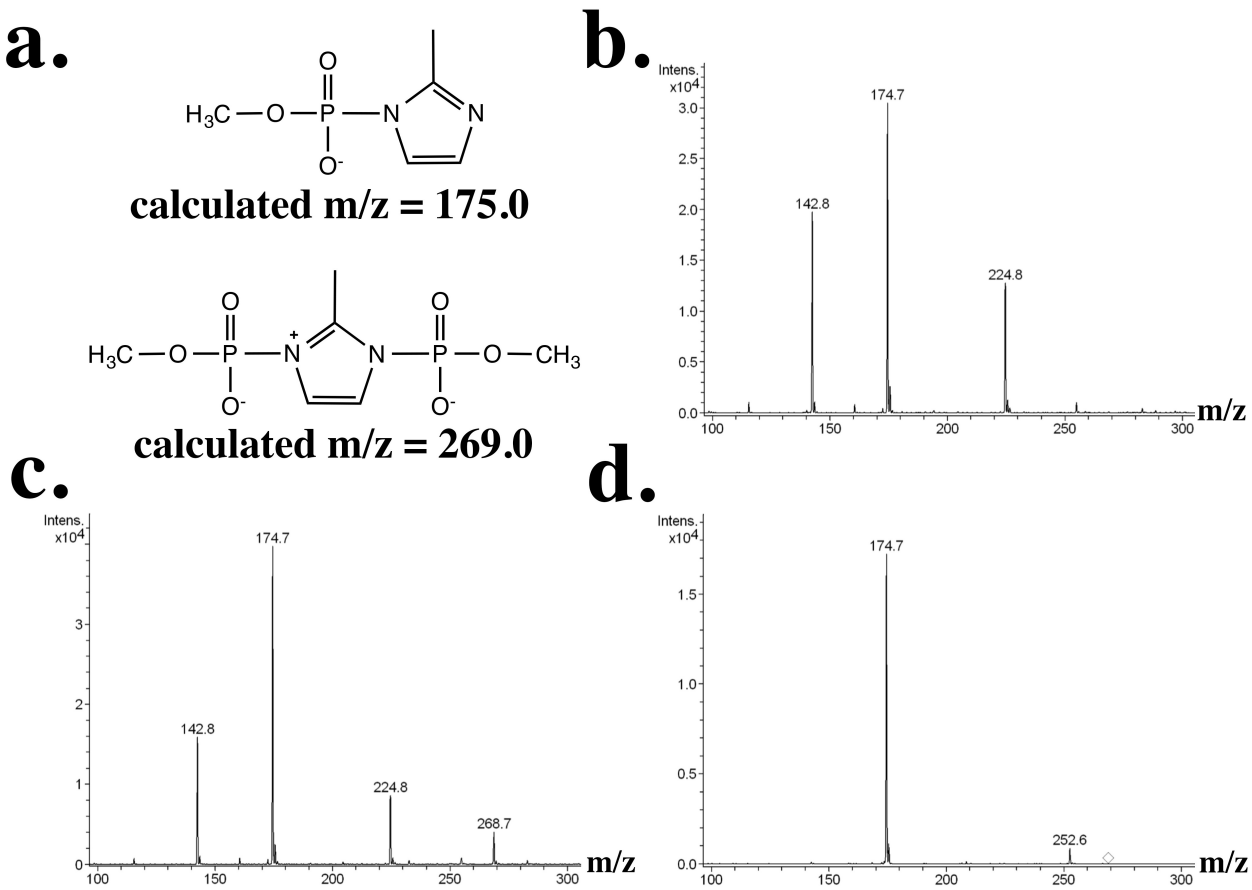

Figure S13. Negative-ion mode ESI-MS analysis of a covalent intermediate formed by incubation of 1-(mono-methyl-phosphoryl)-2-methylimidazole (2-MeImpOMe) at pH 7. (a) 2-MeImpOMe provides a highly simplified analog for the 2-MeImpN monomers. This molecule might also form an imidazolium-bridged dimer with a calculated m/z = 269.0. (b) ESI-MS spectrum of 100 mM 2-MeImpOMe incubated at pH 8.9 for 30 minutes. (c) ESI-MS spectrum of 100 mM 2-MeImpOMe incubated at pH 6.9 for 30 minutes. The new peak at m/z = 268.7 corresponds to the mass of an imidazolium-bridged dimer. (d) MS^2 fragmentation spectrum of m/z = 269, the proposed mass of the proposed dimer. m/z = 174.7 corresponds to 2-MeImpOMe. m/z = 252.6 corresponds to loss of an oxygen atom from the intermediate. m/z = 268.7 corresponds to a mono-methyl-phosphate dimer bonded to 2-methylimidazole. m/z = 142.8 and 224.8 are from impurities.

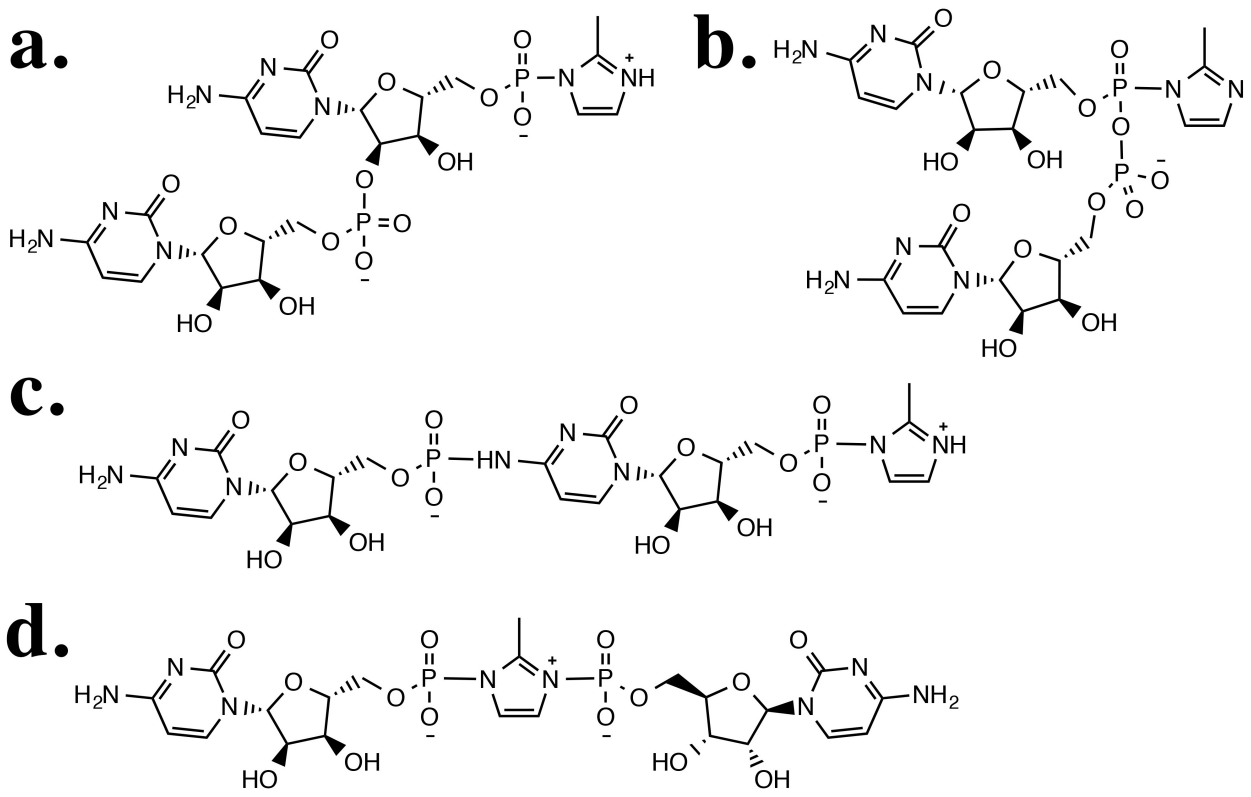

Figure S14. Structural isomers of the di-cytidine intermediate. (a) A phosphodiester-linked dimer with a 5′-(phosphoryl-2-methylimidazole) group. A 2′-5′ linkage is pictured, but a 3′-5′ linkage is also possible. (b) A pyrophosphate-linked dimer with 2-methylimidazole attached to one of the phosphorous atoms. (c) A nucleobase-linked dimer with a bond between phosphate and the exocyclic amine group. Other linkages between phosphate and the nucleobase are possible, such as a bond between phosphate and the aromatic imine nitrogen. (d) The imidazolium-bridged dimer that connects the nucleotides by a 5′-5′ linkage.

	Covalent products formed by self-reaction of:			
Chemical structure of product:	ribonucleotide 2-MeImpC	2′,3′-dideoxy-2-MeImpC	2′,3′-dideoxy-2-MeImpU	2-MeImpOMe
Phosphodiester-linked	✔			
Pyrophosphate-linked	✔			
Nucleobase-linked		✔		
Imidazolium-bridged			✔	✔

Table S1. Summary of analysis by ESI-MS. Formation of a covalent imidazolium bridged intermediate by incubation at neutral pH is observed by NMR for each monomer (Fig. S10–13). However, MS^2 fragmentation suggests that additional isomers are present (Fig. S14). A checkmark indicates that the MS^2 fragmentation spectrum formed by a particular monomer is consistent with the proposed structure. The MS^2 fragmentation spectrum of the dimer formed by ribonucleotide 2-MeImpC (Fig. S10) contains peaks for both the nucleotide monophosphate (m/z = 321.8) and a loss of 2-methylimidazole from the dimer (m/z = 608.9), which is consistent with both the phosphodiester-linked and pyrophosphate-linked dimers. The MS^2 fragmentation spectrum of 2′,3′-dideoxy-2-MeImpC (Fig. S11) contains a peak corresponding to cytosine covalently attached to CMP (m/z = 382.8), which is consistent with a nucleobase-linked dimer. The MS^2 fragmentation spectra of 2′,3′-dideoxy-2-MeImpU and 2-MeImpOMe (Fig. S12–13) do not contain peaks corresponding to the loss of 2-methylimidazole, which is consistent with the imidazolium-bridged structure.

Supplementary text: Consideration of alternative structures for an intermediate in primer extension

The existence of a covalent intermediate in the primer extension reaction was deduced from the effect of monomer stock pH, incubation time, and concentration on k_{obs} (Fig. 1–2). The effect of stock pH and concentration suggested a mechanism involving a chemical reaction between the 2-MeImpN anion and zwitterion to form a dimer (Fig. 3a), or perhaps a longer oligomer. We synthesized material that is enriched for this intermediate, which quickly hydrolyzed into activated monomer and ribonucleotide-5′-monophosphate (Fig. 4). Importantly, using this enriched material in the primer extension reaction resulted in fast reaction rates, indicating that the intermediate is itself a substrate for primer extension.

By NMR spectroscopy, we observed the appearance of new or increased peaks in monomer solutions at pD 7 (compared to pD 8.5 or 4.7) that suggested the formation of a new molecular species (Fig. 3b–c). By ^{31}P NMR, we observed a single new peak at ~ -13.9 ppm for ribonucleotide and 2′,3′-dideoxy-ribonucleotide monomers, indicating that this new peak does not depend on the identity of the nucleobase or result from formation of phosphodiester linkages (Fig. S3–8). Labeling the C2 of 2-methylimidazole with ^{13}C resulted in splitting of the ^{31}P NMR peak at -13.9 ppm, indicating that this peak corresponds to a phosphorous atom bound to 2-methylimidazole (Fig. 5f). ^{13}C NMR of the material enriched for the proposed intermediate, as well as the ^{13}C-isotope labeled monomer, revealed a new triplet that results from coupling of the ^{13}C-C2 of 2-methylimidazole to two equivalent phosphorous atoms (Fig. 4d, 5d). These results indicate that the intermediate formed by incubation at neutral pH is an imidazolium-bridged dinucleotide.

By mass spectrometry, we observed an additional peak when the 2-MeImpC monomer was incubated at neutral pH (Fig. S10). The m/z value of this new peak corresponds to a nucleotide dimer covalently attached to 2-methylimidazole, consistent with the imidazolium-bridged dinucleotide. However, tandem mass spectrometry (MS^2) of the new peak by collision-induced dissociation (CID) resulted in a fragmentation spectrum consistent with the presence of a dimer linked through a phosphodiester bond or a pyrophosphate bond (Fig. S14a–b). Since our NMR spectra strongly indicated the presence of an imidazolium-bridged dinucleotide, we suspect that these mass spectra result from a mixture with at least one structural isomer of the imidazolium-bridged dinucleotide. Using liquid-chromatography mass spectrometry (LC-MS), we have not yet been able to obtain a spectrum uniquely consistent with the structure of an imidazolium-bridged di-cytidine intermediate. Possibly, the highly reactive imidazolium-bridge rearranges during mass spectrometry to form fragments that do not represent the imidazolium-bridged structure.

To eliminate any phosphodiester-linked dimers from the sample, we analyzed 2′,3′-dideoxy ribonucleotide monomers by ESI-MS. These analogs of the RNA monomer lack hydroxyl groups and therefore cannot form phosphodiester bonds. In addition, the lack of ribose hydroxyls prevents possible rearrangement of the imidazolium-bridged dinucleotide to form a phosphodiester-linked dimer. When incubated at pD 7, we observed peaks by NMR that suggest 2′,3′-dideoxy-2-MeImpC and 2′,3′-dideoxy-2-MeImpU also form an imidazolium-bridged dinucleotide (Fig. S7–8). For both of the 2′,3′-dideoxy-ribonucleotide monomers, we observed by ESI-MS the formation of a covalent dimer attached to 2-methylimidazole after incubation at pH 7 (Fig. S11–12). For 2′,3′-dideoxy-2-MeImpC, the CID fragmentation spectrum of the covalent intermediate displayed a peak (m/z = 382.8) diagnostic of an unexpected structural

isomer linked through the nucleobase and phosphate (Fig. S14c). For 2′,3′-dideoxy-2-MeImpU, which cannot form either phosphodiester or nucleobase linked dinucleotides, we did observe a CID fragmentation spectrum of the covalent intermediate that is consistent with an imidazolium-bridged structure. This spectrum does not include a fragment corresponding to loss of 2-methylimidazole, which is expected for the other structural isomers.

To further confirm that formation of an imidazolium bridged covalent dimer does not require the nucleobase or ribose hydroxyls, we analyzed a highly simplified analog of the activated monomer, 1-(mono-methyl-phosphoryl)-2-methylimidazole (2-MeImpOMe) (Fig. S13a). Incubation of 2-MeImpOMe at neutral pH also resulted in observation of a covalent dimer by ESI-MS (Fig. S13b–c). The fragmentation spectrum of this dimer does not include a peak resulting from loss of 2-methylimidazole, which is consistent with the imidazolium-bridged structure (Fig. S13d).

In summary, our analysis by mass spectrometry indicates the presence of several structural isomers with the same mass as the imidazolium-bridged dinucleotide (Table S1). Close inspection of our NMR spectra reveal additional unassigned peaks for compounds present in trace amounts, some of which may correspond to these structural isomers. However, we do not believe that these alternative structural isomers are intermediates in primer extension nor correspond to the ^{31}P NMR signal at -13.9 ppm. Below, we individually discuss the evidence for the existence of each structural isomer as well as the evidence that none can play a role as an intermediate in primer extension.

Phosphodiester-linked dimer: A structural isomer that is likely present in our samples is a ribonucleotide dimer linked through a 3′-5′ or 2′-5′ phosphodiester bond and covalently attached to 2-methylimidazole at the 5′-phosphoryl group (2-MeImpNpN) (Fig. S14a). These molecules

have been previously observed after extended incubation of activated monomers at pH 7.5 and are known catalysts of the primer extension reaction.[1,2] Formation of this dimer can explain some but not all of the kinetic results. For instance, the phosphodiester-linked dimer might be formed during incubation at pH 7, and subsequently increase k_{obs} of the primer extension reaction through a noncovalent catalytic effect. In addition, the MS^2 fragmentation spectrum of the covalent intermediate formed by 2-MeImpC at pH 7 is consistent with this structure (Fig. S10).

However, the phosphodiester-linked dimer is inconsistent with our kinetic analysis of primer extension and NMR characterization of the intermediate formed by incubation at pH 7. First, the 2-MeImpNpN dimer cannot explain why the rate of primer extension can slow down (Fig. 1c, 4c). When a monomer stock solution has been incubated at pH 7 to allow for build up of the intermediate, and is then added to a primer extension reaction, the rate of primer extension is initially very fast but rapidly starts to decrease. This behavior is most simply explained by the rapid hydrolysis of the imidazolium-bridged dinucleotide intermediate. 2-MeImpNpN dimers have been previously synthesized without observation of instability.[1] In fact, these activated oligomers can catalyze primer extension over the course of days. This is in contrast to the rapid hydrolysis of the imidazolium-bridged intermediate (Fig. 4b), which explains the burst kinetics as previously described.[3]

Second, the NMR characterization of the imidazolium-bridged intermediate cannot correspond to the 2-MeImpCpC structure. By ^{31}P NMR, the -13.9 ppm peak indicative of the intermediate is observed in samples of 2′,3′-dideoxy monomers, which excludes assignment of this peak to a phosphodiester linkage. Typically, phosphodiester linkages are observed in the region of -2 to -5 ppm, which is not a predominant feature of the ^{31}P NMR spectrum of the partially purified imidazolium-bridged intermediate before it decays. Importantly, the 2-

MeImpCpC structure cannot explain the results obtained using the ^{13}C isotope labeled 2-MeImpC monomer (Fig. 5). The triplet observed by ^{13}C NMR indicates that the C2 carbon atom of 2-methylimidazole is split by two phosphorous atoms, which is not predicted by the phosphodiester-linked dimer.

Pyrophosphate-linked intermediate: A different potential structure for the intermediate involves a 5′-5′ pyrophosphate linkage between two nucleotides with 2-methylimidazole bound to one of the phosphate groups (Fig. S14b). A pyrophosphate-linked dinucleotide could potentially form through the nucleophilic attack of a phosphate oxygen atom on the phosphate group of another activated monomer, releasing free 2-methylimidazole as a leaving group. The pyrophosphate-linked dimer is consistent with the fragmentation spectrum obtained for the m/z = 691 peak in samples of 2-MeImpC incubated at pH 7 (Fig. S10).

However, the pyrophosphate-linked dimer is not a plausible intermediate in primer extension and does correspond to our characterization of the intermediate by NMR spectroscopy. For the pyrophosphate-linked dimer to be an intermediate of the primer extension reaction, the pyrophosphate linkage would have to be broken to extend the primer by one nucleotide, a reaction that is expected to be quite slow. Furthermore, the pyrophosphate dimer should hydrolyze to form CppC and free 2-methylimidazole. However, we do not observe formation of CppC during the decay of the di-cytidine intermediate (Fig. 4b).

Second, the pyrophosphate-linked dimer cannot correspond to the partially-purified intermediate formed at neutral pH. Based upon the structure of the pyrophosphate-linked dimer, we expect the nonequivalent atoms to couple with each other, resulting in two doublets by ^{31}P NMR. Instead, we observe that the intermediate results in a singlet peak at -13.9 ppm, indicating that this peak cannot be assigned to a pyrophosphate-linked dinucleotide with 2-methylimidazole

bound to one of the phosphates unless the imidazole exchanges rapidly between phosphates. In addition, the pyrophosphate link cannot explain the triplet observed by ^{13}C NMR for the partially purified material and the ^{13}C isotope labeled monomer (Fig. 4d, 5d).

Nucleobase-linked dimer: An alternative potential structure for an intermediate of primer extension is a dimer linked through a bond between the nucleobase and phosphate (Fig. S14c). We began to consider this possible structure after observing the MS^2 fragmentation spectrum of the intermediate formed by 2′,3′-dideoxy-2-MeImpC (Fig. S11). This spectrum contains a diagnostic peak (m/z = 382.8) that can only be explained by a nucleobase-linked dimer. Also, we expect that a linkage between the nucleobase and phosphate would be reactive, consistent with a potential role as an intermediate in primer extension.

However, the nucleobase-linked dimer cannot explain all of the kinetic results. First, the effect of 2-MeImpN monomer stock pH on the rate of primer extension is largely independent of the identity of the nucleobase (Fig. 2a). The three nucleotides all have the same optimal stock pH. If the intermediate responsible for the effect of monomer stock pH on the initial rate of primer extension involved bonds with the nucleobase, then we expect that the optimal stock pH would depend on the identity of the nucleobase. Second, the nucleobase-linked dimer cannot account for the catalytic effect of 2-methylimidazole. Previous results indicate that downstream oligomers catalyze primer extension only if they are activated by 2-methylimidazole. If catalysis involved a nucleobase-linked intermediate, then the catalytic effect would require a 2-methylimidazole group on the incoming monomer, but not the oligomer. Finally, the nucleobase-phosphate linkage would prevent this intermediate from binding the template as a dimer. This presents difficulties for how this minor component of the monomer stock solution could

outcompete the other abundant nucleotides for binding to the template and reacting with the primer.

In addition, our NMR characterization of the imidazolium-bridged intermediate cannot correspond to the nucleobase-linked dimer. The chemical shift of the -13.9 ppm peak indicative of the intermediate is largely independent of the identity of the nucleobase (Fig. S3–6). If the intermediate contained a linkage between phosphate and the nucleobase, we would expect the chemical shift to depend on the identity of the nucleobase, but this is not observed for the peak at -13.9 ppm in ^{31}P NMR spectra. In addition, the -13.9 ppm peak is split by the ^{13}C-isotope label, which excludes assignment of this peak to a phosphate bound to the nucleobase. The structure of the nucleobase-linked dimer also does not explain the triplet observed by ^{13}C NMR using the ^{13}C-isotope labeled 2-MeImpC monomer (Fig. 5).

Imidazolium-bridged dimer: Our proposed structure for an intermediate of primer extension is an imidazolium-bridged dinucleotide that contains a 5′-5′ linkage (Fig. S14d). This structure is supported by our kinetic analysis of the primer extension reaction, the decay of the intermediate, NMR spectroscopy, and mass spectrometry. First, our kinetic analysis indicated that incubation of the stock of the monomer at pH 7 resulted in the fastest rates of primer extension. This pH corresponds to the pK_a of the 2-methylimidazole group of 2-MeImpG, suggesting that the mechanism of intermediate formation involves a reaction between the zwitterionic and anionic forms of the monomer (Fig. 3a). The resulting imidazolium-bridged dimer would be a highly reactive substrate of the primer extension reaction that might bind as a dimer (Fig. 6). Although other structural isomers are possible, the imidazolium-bridged dinucleotide explains that effect of monomer stock pH on the initial rate of primer extension.

The structure of an imidazolium-bridged dimer is also consistent with our analysis of the partially purified material (Fig. 4). This material is a substrate of the primer extension reaction (Fig. 4c), consistent with our expectation from the effect of stock pH on the rate of primer extension. As a substrate, this intermediate successfully accounts for previously described features of the primer extension reaction, such as the difficulty of copying the last nucleotide of a template and the catalytic effect of downstream activated oligomer. In addition, the hydrolysis of the di-cytidine intermediate to form CMP and 2-MeImpC is expected for the imidazolium-bridged structure.

The imidazolium-bridged dimer is the only chemical structure that is consistent with the results obtained by NMR spectroscopy. Incubation of multiple monomers at pD 7 results in the increase of a singlet peak at -13.9 ppm in ^{31}P NMR spectra (Fig. S3–8), which we subsequently assigned to the imidazolium-bridged dinucelotide (Fig. 5). The single phosphorous resonance of the intermediate supports the symmetric structure of the proposed intermediate. The structure of the imidazolium-bridged dimer also correctly predicted that the C2 carbon of 2-methylimidazole would be split by two phosphorous atoms, resulting in a triplet signal. This prediction was confirmed in our analysis of the intermediate-enriched material and the ^{13}C isotope labeled monomer (Fig. 4d, 5d). In addition, this structure correctly predicted that the -13.9 ppm ^{31}P NMR peak of the intermediate would be split by the ^{13}C isotope label (Fig. 5f).

Our mass spectrometry results have provided both supportive as well as some opposing evidence for the imidazolium-bridged dinucleotide. When activated monomer is incubated at neutral pH, we observed an additional prominent peak corresponding to a dinucleotide bound to 2-methylimidazole (Fig. S10–13), which is consistent with our expected mass of the imidazolium-bridged dinucleotide. For 2′,3′-dideoxy-2-MeImpU and the analog 2-MeImpOMe

(Fig. S12–13), we obtained simple MS^2 fragmentation spectra of the dimers formed at neutral pH that are consistent with an imidazolium-bridged structure. However, the imidazolium-bridged structure is inconsistent with the MS^2 spectra of the dimers formed by 2-MeImpC and 2′,3′-dideoxy-2-MeImpC (Fig. S10–11). These spectra both contain a peak resulting from loss of 2-methylimidazole, which indicates that the structure of this fragment cannot be linked through an imidazolium bridge. We currently suspect that the MS^2 spectra from our analysis of 2-MeImpC and 2′,3′-dideoxy-2-MeImpC may result from the presence of at least one structural isomer in addition to the imidazolium-bridged dinucleotide. However, we have not been able to separate these putative structural isomers from the imidazolium-bridged dinucleotide using liquid chromatography, which might be expected due to the similar physical properties of structural isomers. We are also considering that the highly reactive imidazolium-bridged dinucleotide rearranges during mass spectrometry with either the ribose hydroxyls or the exocyclic amine of a nucleobase. Although this hypothesis is *post hoc*, the possibility of rearrangement during mass spectrometry cannot be excluded unless we obtain mass spectra solely consistent with the imidazolium-bridged dinucleotide for all monomers.

Finally, we note that the confirmation of an imidazolium-bridged structure through tandem mass spectrometry is difficult due to the symmetry of this molecule. While certain fragments can be diagnostic of a particular structure, all of the fragments produced by the imidazolium-bridged dimer can be explained by the presence of one or more alternative structural isomers. This means that no fragment of the intermediate can be diagnostic of the imidazolium-bridged structure. The only fragmentation spectra that can support the imidazolium-bridged structure would have an absence of diagnostic fragments for other structures. This result was obtained for our analysis of both 2′,3′-dideoxy-2-MeImpU and 2-MeImpOMe (Table S1).

Therefore, we do not expect that tandem mass spectrometry can provide definitive evidence for the imidazolium-bridged structure.

Supplementary methods:

Synthesis of CppC. Di-cytidine-5′,5′-pyrophosphate (Fig. S9) was synthesized through a previously published protocol.[4]

Synthesis of 1-(mono-methyl-phosphoryl)-2-methylimidazole. 50 μL of phosphorous(V) oxychloride is slowly added to a room temperature mixture of 3 mL methanol, 450 μL triethylamine, and 25 mg 2-methylimidazole under non-anhydrous conditions. The reaction is quenched after 2 hours by drop-wise addition to 7 mL water and stirred 15 minutes. The resulting solution is briefly placed on a rotary evaporator to remove the methanol, followed by column chromatography and lyophilization as described for the purification of 2-MeImpN monomers.

Supplementary references:

1. Prywes, N.; Blain, J.C.; Del Frate, F.; Szostak, J.W. *Elife* **2016**, *5*, e17756.

2. Kanavarioti, A. *Orig Life Evol Biosph.* **1997**, *27*, 357.

3. Kervio, E.; Sosson, M.; Richert, C. *Nucleic Acids Res*. **2016**, *44,* 5504**.**

4. Kanavarioti, A.; Lu, J.; Rosenbach, M.T.; Hurley, T.B. *Tetrahedron Lett*. **1991**, *32*, 6065.